Title: Les langues sur le web

Author: Marie Lebert

Release Date: April 19, 2010 [EBook #32065]

Language: French

Character set encoding: ISO-8859-1

Produced by Al Haines

Les langues sur le web

Marie Lebert

NEF, Université de Toronto, 2010

Ce livre est dédié à toutes les personnes
interviewées de par le monde
sur ce vaste sujet.

On dit souvent que l'internet abolit le temps, les distances et les frontières, mais qu'en est-il des langues? Comme l'écrit si bien Maria Victoria Marinetti, professeure d'espagnol en entreprise et traductrice, «il est très important de pouvoir communiquer en différentes langues. Je dirais même que c'est obligatoire, car l'information donnée sur l'internet est à destination du monde entier, alors pourquoi ne l'aurions-nous pas dans notre propre langue ou dans la langue que nous souhaitons utiliser? Information mondiale, mais pas de vaste choix dans les langues, ce serait contradictoire, pas vrai?» Tel est le sujet de ce livre. Il s'agit de la version revue et actualisée de *L'internet et les langues*, publié en 2009.

***

Marie Lebert, chercheuse et journaliste, s'intéresse aux technologies dans le monde des médias et des langues. Elle est l'auteure des livres suivants: *Le livre, de l'imprimé au numérique* (NEF, 2010), *L'internet et les langues* (NEF, 2009) et *Le multilinguisme sur le web* (NEF, 1999). Ses livres sont publiés par le NEF (Net des études françaises), Université de Toronto, et sont librement disponibles sur le site du NEF <www.etudes-francaises.net> et dans le Projet Gutenberg <www.gutenberg.org>, dans divers formats permettant leur lecture sur tout appareil électronique.

Table

Introduction

On dit souvent que l'internet abolit le temps, les distances et les frontières, mais qu'en est-il des langues?

Des «nations de langues»
«Comme l’internet n’a pas de frontières nationales, les internautes s’organisent selon d’autres critères propres au médium. En termes de multilinguisme, vous avez des communautés virtuelles, par exemple ce que j’appelle les "nations de langues", tous ces internautes qu’on peut regrouper selon leur langue maternelle quel que soit leur lieu géographique. Ainsi la nation de la langue espagnole inclut non seulement les internautes d’Espagne et d’Amérique latine, mais aussi tous les Hispanophones vivant aux États-Unis, ou encore ceux qui parlent espagnol au Maroc.» (Randy Hobler, consultant en marketing internet chez Globalink, société de produits et services de traduction, septembre 1998)

La «démocratie linguistique»
«Dans un rapport de l'UNESCO du début des années 1950, l'enseignement dispensé dans sa langue maternelle était considéré comme un droit fondamental de l'enfant. La possibilité de naviguer sur l'internet dans sa langue maternelle pourrait bien être son équivalent à l'Âge de l'Information. Si l'internet doit vraiment devenir le réseau mondial qu'on nous promet, tous les usagers devraient y avoir accès sans problème de langue. Considérer l'internet comme la chasse gardée de ceux qui, par accident historique, nécessité pratique ou privilège politique, connaissent l'anglais, est injuste à l'égard de ceux qui ne connaissent pas cette langue.» (Brian King, directeur du WorldWide Language Institute, septembre 1998)

Un médium pour le monde
«Il est très important de pouvoir communiquer en différentes langues. Je dirais même que c’est obligatoire, car l’information donnée sur l'internet est à destination du monde entier, alors pourquoi ne l’aurions-nous pas dans notre propre langue ou dans la langue que nous souhaitons utiliser? Information mondiale, mais pas de vaste choix dans les langues, ce serait contradictoire, pas vrai?» (Maria Victoria Marinetti, professeure d’espagnol en entreprise et traductrice, août 1999)

De bons logiciels...
«Quand la qualité des logiciels sera suffisante pour que les gens puissent converser par écrit et par oral sur le web en temps réel dans différentes langues, nous verrons tout un monde s'ouvrir à nous. Les scientifiques, les hommes politiques, les hommes d'affaires et bien d'autres groupes seront à même de communiquer immédiatement entre eux sans l'intermédiaire de médiateurs ou traducteurs.» (Tim McKenna, écrivain et philosophe, octobre 2000)

...dans toutes les langues
«Les recherches sur la traduction automatique devraient permettre une traduction automatique dans les langues souhaitées, mais avec des applications pour toutes les langues et non les seules dominantes (ex.: diffusion de documents en japonais, si l’émetteur est de langue japonaise, et lecture en breton, si le récepteur est de langue bretonne). Il y a donc beaucoup de travaux à faire dans le domaine de la traduction

automatique et écrite de toutes les langues.» (Pierre-Noël Favennec, expert à la direction scientifique de France Télécom R&D, février 2001)

Des dictionnaires de langues en ligne
«J'ai débuté WordReference.com en 1999 pour procurer des dictionnaires bilingues gratuits en ligne et d'autres outils pour tous sur l'internet. Depuis, le site s'est progressivement développé pour devenir l'un des sites de dictionnaires en ligne les plus utilisés, et le principal dictionnaire en ligne pour les paires de langues anglais-espagnol, anglais-français, anglais-italien, espagnol-français et espagnol-portugais. Ce site est toujours classé sans interruption parmi les 500 sites les plus visités du web. Aujourd'hui, je suis heureux de continuer à améliorer ces dictionnaires, les autres outils linguistiques du site et les forums de langues. J'ai vraiment plaisir à créer de nouvelles fonctionnalités pour rendre ce site de plus en plus utile.» (Michael Kellogg, fondateur de WordReference.com, mars 2010)

***

Ce livre est issu des multiples liens tissés sur le Net des études françaises (NEF), fondé en mai 2000 par Russon Wooldridge, professeur à l'Université de Toronto (Canada). Sauf indication contraire, les citations sont des extraits des Entretiens du NEF <http://www.etudes-francaises.net/entretiens/> et des entretiens qui ont suivi pour les actualiser et les compléter.

L'Ethnologue

Un catalogue exhaustif des langues
Une œuvre de longue haleine
La codification des langues

Un catalogue exhaustif des langues

Le nombre de langues vivantes parlées sur notre planète s'élève très exactement à 6.909, un décompte datant de 2009. Ces langues vivantes sont toutes recensées dans L'*Ethnologue: Languages of the World*, catalogue exhaustif à la fois par la masse d'informations glanées et par la qualité de ces informations, vérifiées une à une sur le terrain dans le monde entier par un réseau composé de milliers de linguistes.

Travaillant au sein de l'organisme SIL International (SIL signifiant: Summer Institute of Linguistics), l'équipe de l'Ethnologue insiste à juste titre sur la minutie avec laquelle les informations sont collectées puis vérifiées. Une petite équipe de chercheurs basée à Dallas, dans le Texas, est elle-même relayée par des équipes nationales et linguistiques présentes sur tous les continents.

Barbara Grimes, directrice de publication entre 1971 et 2000 (8e-14e éditions), relate en janvier 2000 : « Il s’agit d’un catalogue des langues dans le monde, avec des informations sur les pays où elles sont parlées, une estimation du nombre de personnes qui les parlent, la famille linguistique à laquelle elles appartiennent, les autres termes utilisés pour ces langues, les noms de dialectes, diverses informations socio-linguistiques et démographiques, les dates des Bibles publiées, un index des noms de langues, un index des familles linguistiques et enfin des cartes géographiques pour les langues.»

L'index des noms de langues (Ethnologue Name Index) donne la liste des noms de langues et de dialectes ainsi que leurs synonymes, et l'index des familles linguistiques (Ethnologue Language Family Index) organise les langues selon leurs familles linguistiques.

Lors d'un entretien précédent, en août 1998, Barbara Grimes explique aussi que, si la version web est utile, la version imprimée l'est encore plus, en langue anglaise pour toucher un large public: «Nous avons eu des demandes nous demandant l'accès à l'*Ethnologue* dans plusieurs autres langues, mais nous n'avons pas le personnel ni les fonds pour la traduction ou l'actualisation, indispensable puisque notre site est constamment mis à jour. L'internet est (...) un moyen commode pour mettre notre documentation à la disposition d'une audience plus large que celle de l'*Ethnologue* imprimé. D'un autre côté, l'*Ethnologue* sur l'internet n'atteint en fait qu'une audience limitée disposant d'ordinateurs. Or, dans les personnes que nous souhaitons atteindre, nombreuses sont celles qui ne disposent pas d'ordinateurs. Je pense particulièrement aux habitants du dit "Tiers-monde".»

Une nouvelle édition de l'*Ethnologue* est publiée environ tous les quatre ans. L'édition la plus récente, la 16e, date de 2009, avec une version imprimée payante et une version web gratuite... comme d'habitude depuis que le web existe. Les éditions précédentes

proposaient aussi une version CD-ROM payante, pendant quelques années, avant que l'internet à débit rapide ne devienne monnaie courante.

Une œuvre de longue haleine

Contrairement à ce que certains croient, cette vaste entreprise n'est née ni avec l'internet (1974) ni avec le web (1993), même si sa version web a beaucoup contribué à la faire connaître. Il s'agit d'un projet de recherche actif depuis plus de cinquante ans.

L'*Ethnologue* est fondé en 1951 par Richard Pittman, qui est le directeur de publication des sept premières éditions, de 1951 à 1969. À l'origine, son but est de partager les informations glanées sur les langues minoritaires avec ses collègues de SIL International et d'autres linguistes. L'Ethnologue débute comme un catalogue de langues minoritaires avant de prendre de l'ampleur au fil des ans.

Barbara Grimes est la directrice de publication des sept éditions suivantes, de la 8e à la 14e, entre 1971 à 2000. À partir de 1971, le champ des informations s'élargit pour ne plus concerner seulement les langues minoritaires, mais pour englober aussi peu à peu toutes les langues vivantes connues dans le monde. Entre 1967 et 1973, Barbara Grimes révise d'abord en profondeur toutes les informations disponibles sur les langues d'Afrique, des Amériques, du Pacifique et de quelques pays d'Asie. Pendant ses années comme directrice de publication, entre 1971 et 2000, le nombre de langues recensées passe de 4.493 à 6.809, et les informations disponibles pour chaque langue sont elles-mêmes plus complètes, si bien que l'encyclopédie triple de taille.

En 2000, Raymond Gordon Jr. devient le troisième directeur de publication de l'*Ethnologue* et produit la 15e édition (2005). Il cède ensuite la place à Paul Lewis, aidé de Conrad Hurd pour la gestion de la base de données et de Raymond Gordon pour la direction des équipes nationales et linguistiques sur le terrain.

Dans l'introduction de la 16e édition (2009), on peut lire ceci : «La manière dont chacun choisit de définir une langue dépend des motifs qu'on a d'identifier cette langue comme étant distincte d'une autre. Certains basent la définition d'une langue sur des raisons purement linguistiques. D'autres reconnaissent la nécessité de prendre également en compte des facteurs sociaux, culturels ou politiques. En outre, les locuteurs d'une langue donnée ont souvent leurs propres critères sur l'appropriation d'une langue comme étant la leur. Ces critères sont souvent bien davantage liés à des questions de patrimoine et d'identité qu'aux traits linguistiques de la langue ou des langues en question.»

La codification des langues

Comme expliqué dans l'introduction de la 16e édition (2009), une caractéristique de la base de données de l'*Ethnologue* est un système de codification des langues au moyen de trois lettres (par exemple FRA pour la France), ce depuis sa création. Cette codification est incluse dans l'encyclopédie elle-même à partir de la 10e édition (1984).

L'Organisation internationale de normalisation (ISO) adopte pour sa part la norme ISO 639-2 en 1998. Cette norme, qui codifie également chaque langue au moyen de trois lettres, est une convergence de la norme ISO 639-1 (un ancien standard de deux lettres pour chaque langue, adopté en 1988) et de la norme ANSI Z39.53 (les codes de langues

MARC, trois lettres pour chaque langue, développés par les bibliothèques américaines et adoptés en tant que norme nationale en 1987).

La norme ISO 639-2 (1998) devient vite insuffisante puisqu'elle codifie moins de 400 langues individuelles. En 2002, l'ISO invite donc SIL International à établir une nouvelle norme permettant d'harmoniser les codes utilisés dans l'*Ethnologue* avec les codes de la norme ISO, en y intégrant aussi les codes des langues mortes utilisés par la Linguist List, une grande liste de diffusion à destination des linguistes.

Le résultat, officiellement approuvé en 2006 et publié en 2007, est la nouvelle norme ISO 639-3, qui attribue un code de trois lettres à près de 7.500 langues. SIL International est également désigné comme l'organisme responsable de l'enregistrement de nouvelles langues pour l'ISO 639-3, et gère donc le cycle annuel des modifications et des mises à jour sur une page de son propre site.

De l'ASCII à l'Unicode

L'ASCII

Communiquer dans plusieurs langues implique d’avoir des systèmes d'encodage adaptés à nos alphabets ou idéogrammes respectifs.

Le premier système d'encodage informatique est l’ASCII (American standard code for information interchange). Publié en 1968 aux États-Unis par l’American National Standards Institute (ANSI), avec actualisation en 1977 et 1986, l'ASCII est un code standard de 128 caractères traduits en langage binaire sur sept bits (A est traduit par «1000001», B est traduit par «1000010», etc.). Les 128 caractères comprennent 33 caractères de contrôle (qui ne représentent donc pas de symbole écrit) et 95 caractères imprimables: les 26 lettres sans accent en majuscules (A-Z) et minuscules (a-z), les chiffres, les signes de ponctuation et quelques caractères spéciaux, le tout correspondant aux touches du clavier anglophone.

L'ASCII permet uniquement la lecture de l’anglais et du latin. Il ne permet pas de prendre en compte les lettres accentuées présentes dans bon nombre de langues européennes, et à plus forte raison les langues non alphabétiques (chinois, japonais, coréen, etc.). Ceci ne pose pas de problème majeur les premières années, tant que l’échange de fichiers électroniques se limite essentiellement à l’Amérique du Nord. Mais le multilinguisme devient bientôt une nécessité vitale. Des variantes de l’ASCII (norme ISO-8859 ou ISO-Latin) prennent en compte les caractères accentués de quelques langues européennes. La variante pour le français, par exemple, est définie par la norme ISO-8859-1 (ISO-Latin-1).

L'Unicode

Avec le développement du web, l’échange des données s’internationalise de plus en plus. On ne peut plus se limiter à l’utilisation de l’anglais et de quelques langues européennes, traduites par un système d’encodage datant de 1968. De plus, le passage de l’ASCII original à ses différentes extensions devient vite un véritable casse-tête, y compris au sein de l’Union européenne, les problèmes étant entre autres la multiplication des variantes, la corruption des données dans les échanges informatiques ou encore l’incompatibilité des systèmes, les pages ne pouvant être affichées que dans une seule langue à la fois.

Olivier Gainon, fondateur de CyLibris et pionnier de l’édition électronique littéraire, écrit en décembre 2000: «Il faut que le réseau respecte les lettres accentuées, les lettres spécifiques, etc. Je crois très important que les futurs protocoles permettent une transmission parfaite de ces aspects - ce qui n’est pas forcément simple (dans les futures évolutions de l’HTML ou des protocoles IP, etc.). Donc il faut que chacun puisse se sentir à l’aise avec l’internet et que ce ne soit pas simplement réservé à des (plus ou moins) anglophones. Il est anormal aujourd’hui que la transmission d’accents puisse poser problème dans les courriers électroniques. La première démarche me semble

donc une démarche technique. Si on arrive à faire cela, le reste en découle: la représentation des langues se fera en fonction du nombre de connectés, et il faudra envisager à terme des moteurs de recherche multilingues.»

Publié pour la première fois en janvier 1991, l'Unicode est un système d'encodage universel sur 16 bits spécifiant un nombre unique pour chaque caractère. Ce nombre est lisible quels que soient la plateforme, le logiciel et la langue utilisés. L'Unicode peut traiter 65.000 caractères uniques et prendre en compte tous les systèmes d'écriture de la planète. A la grande satisfaction des linguistes, il remplace progressivement l'ASCII, avec des variantes UTF-8, UTF-16 et UTF-32 (UTF: Unicode transformation format) en fonction du nombre de bits utilisés. Il devient une composante des spécifications du World Wide Web Consortium (W3C), l'organisme international chargé du développement du web.

L'utilisation de l'Unicode se généralise à partir de 1998, par exemple pour les fichiers texte sous plateforme Windows (Windows NT, Windows 2000, Windows XP et versions suivantes), qui étaient jusque-là en ASCII.

Mais l'Unicode ne peut résoudre tous les problèmes, comme le souligne en juin 2000 Luc Dall'Armellina, co-auteur et webmestre d'oVosite, un espace d'écriture hypermédia: «Les systèmes d'exploitation se dotent peu à peu des kits de langues et bientôt peut-être de polices de caractères Unicode à même de représenter toutes les langues du monde; reste que chaque application, du traitement de texte au navigateur web, emboîte ce pas. Les difficultés sont immenses: notre clavier avec ses ± 250 touches avoue ses manques dès lors qu'il faille saisir des Katakana ou Hiragana japonais, pire encore avec la langue chinoise. La grande variété des systèmes d'écriture de par le monde et le nombre de leurs signes font barrage. Mais les écueils culturels ne sont pas moins importants, liés aux codes et modalités de représentation propres à chaque culture ou ethnie.»

Patrick Rebollar, professeur de littérature française au Japon et modérateur de la liste de diffusion LITOR (littérature et ordinateur), donne son sentiment en janvier 2000: «Il s'agit d'abord d'un problème logiciel. Comme on le voit avec Netscape ou Internet Explorer, la possibilité d'affichage multilingue existe. La compatibilité entre ces logiciels et les autres (de la suite Office de Microsoft, par exemple) n'est cependant pas acquise. L'adoption de la table Unicode devrait résoudre une grande partie des problèmes, mais il faut pour cela réécrire la plupart des logiciels, ce à quoi les producteurs de logiciels rechignent du fait de la dépense, pour une rentabilité qui n'est pas évidente car ces logiciels entièrement multilingues intéressent moins de clients que les logiciels de navigation.»

ASCII et/ou Unicode

Le Projet Gutenberg est fondé dès 1971 par Michael Hart pour numériser les œuvres littéraires et les mettre gratuitement à la disposition de tous.

Les œuvres sont numérisées en mode texte, en utilisant l'ASCII original sur sept bits ou, pour les langues avec accents, l'ASCII sur huit bits pour une prise en compte des caractères accentués. Mais, même dans ce cas, le Projet Gutenberg propose systématiquement en complément une version ASCII sur sept bits sans accents. Sauf,

bien entendu, dans le cas de langues non encodables en ASCII, comme le chinois, qui est encodé au format Big-5 pour le chinois traditionnel.

Surnommé à juste raison «le plus petit dénominateur commun», l'ASCII sur sept bits est le seul format compatible avec 99% des machines et des logiciels, et pouvant être converti dans de nombreux autres formats. Il sera toujours utilisé quand d’autres formats auront disparu. Il est l’assurance que les collections ne deviendront jamais obsolètes, et survivront aux changements technologiques des prochaines décennies ou même des prochains siècles. Il n'existe pas d'autre standard aussi largement utilisé, y compris l'Unicode, système d'encodage universel créé en 1991. Ce jusqu'en 2008, date à laquelle les deux systèmes d'encodage sont également représentés sur le web.

Le Projet Gutenberg propose certains livres dans d’autres formats que l'ASCII, notamment dans les trois formats répandus que sont les formats HTML, XML et RTF. Des fichiers Unicode sont également présents et remplacent peu à peu les fichiers ASCII sur huit bits pour les langues avec accents comme le français. De plus, tout format proposé par tel ou tel volontaire est généralement accepté (PDF, LIT, TeX et beaucoup d'autres), dans la mesure où un fichier ASCII est également présent.

Les collections du Projet Gutenberg sont d'abord essentiellement anglophones, puisque ce projet est basé aux États-Unis et qu'il sert en priorité la communauté anglophone nationale et internationale. En octobre 1997, Michael Hart annonce son intention d'intensifier la production de livres dans des langues autres que l'anglais. Début 1998, le catalogue comprend quelques œuvres en allemand, en espagnol, en français (dix titres), en italien et en latin. En juillet 1999, Michael écrit: «J'introduis une nouvelle langue par mois maintenant, et je vais poursuivre cette politique aussi longtemps que possible.»

Le multilinguisme devient ensuite l'une des priorités du Projet Gutenberg, tout comme l'internationalisation, avec le Project Gutenberg Australia (créé en août 2001), le Projet Gutenberg Europe (créé en janvier 2004), le Project Gutenberg Canada (créé en juillet 2007), et d'autres Projets Gutenberg à venir dans divers pays, notamment au Portugal.

Dans le Projet Gutenberg original, 25 langues sont représentées en janvier 2004, 42 langues en juillet 2005, dont le sanscrit et les langues mayas, et 50 langues en décembre 2006. À cette date, les langues principales sont l’anglais, le français, l’allemand, le finnois, le hollandais, l’espagnol, l’italien, le chinois, le portugais et le tagalog.

Dès ses débuts en janvier 2004, le Projet Gutenberg Europe propose un site multilingue pour gérer la relecture partagée entre les volontaires, afin de prendre en compte les principales langues nationales. En avril 2004, grâce à des traducteurs volontaires, le site est disponible en douze langues. L'objectif à moyen terme est un site en soixante langues, et donc soixante équipes linguistiques, avec prise en compte de toutes les langues européennes. Le Projet Gutenberg Europe utilise l'Unicode et non l'ASCII, pour pouvoir traiter des livres dans un grand nombre de langues.

Des «nations de langues»

Consultant en marketing internet chez Globalink, une société de produits et services de traduction, Randy Hobler écrit en septembre 1998: «Comme l'internet n'a pas de frontières nationales, les internautes s'organisent selon d'autres critères propres au médium. En termes de multilinguisme, vous avez des communautés virtuelles, par exemple ce que j'appelle les "nations de langues", tous ces internautes qu'on peut regrouper selon leur langue maternelle quel que soit leur lieu géographique. Ainsi la nation de la langue espagnole inclut non seulement les internautes d'Espagne et d'Amérique latine, mais aussi tous les Hispanophones vivant aux États-Unis, ou encore ceux qui parlent espagnol au Maroc.»

Si Randy donne l'exemple d'une «nation de langue» hispanophone répartie sur trois continents, la même remarque vaut pour la Francophonie, une communauté de langue française présente sur cinq continents, ou encore la Créolophonie, puisque le créole est parlé non seulement dans les Caraïbes mais aussi à Paris, Montréal et New York.

À ses débuts, l'internet est anglophone à pratiquement 100%, ce qui s'explique par le fait qu'il débute aux États-Unis en tant que réseau reliant les agences gouvernementales et les universités, suite à la création du protocole TCP/IP en 1974 par Vinton Cerf et Robert Kahn. Après la création du World Wide Web en 1990 par Tim Berners-Lee au Centre européen pour la recherche nucléaire (CERN) à Genève (Suisse) et le lancement en novembre 1993 du premier navigateur Mosaic, ancêtre de Netscape, l'internet se développe rapidement, d'abord aux États-Unis grâce aux investissements considérables du gouvernement, puis au Canada, puis dans le monde entier.

Après avoir été anglophone à pratiquement 100%, l'internet est encore anglophone à plus de 80% en 1998, un pourcentage qui s'explique par trois facteurs: (a) la création d'un grand nombre de sites web émanant des États-Unis, du Canada et du Royaume-Uni; (b) une proportion d'usagers particulièrement forte en Amérique du Nord par rapport au reste du monde, les ordinateurs étant bien meilleur marché qu'ailleurs, tout comme la connexion à l'internet sous forme de forfait mensuel à prix modique; (c) l'usage de l'anglais en tant que principale langue d'échange internationale.

Dans plusieurs pays d'Europe, par exemple, cette connexion est d'abord tarifée à la durée, avec un tarif de jour et un tarif de nuit moins élevé. Les usagers passent donc beaucoup moins de temps sur l'internet qu'ils ne le souhaiteraient, et choisissent souvent de surfer la nuit pour éviter les factures trop élevées. Fin 1998 et début 1999, des mouvements de grève sont lancés en France, en Italie et en Allemagne pour faire pression sur les sociétés prestataires afin qu'elles baissent leurs prix et qu'elles proposent des forfaits internet, avec gain de cause dans les mois qui suivent.

En 1997, Babel, initiative conjointe d'Alis Technologies et de l'Internet Society, mène la première étude sur la répartition des langues sur l'internet. Datée de juin 1997, le *Palmarès des langues de la toile* donne les pourcentages de 82,3% pour l'anglais, 4% pour l'allemand, 1,6% pour le japonais, 1,5% pour le français, 1,1% pour l'espagnol, 1,1% pour le suédois et 1% pour l'italien.

Dans un article publié le 21 juillet 1998 par *ZDNN (ZDNetwork News)*, Martha Stone, journaliste, précise: «Cette année, le nombre de nouveaux sites non anglophones va probablement dépasser celui de nouveaux sites anglophones, et le monde cyber est en train de véritablement devenir une toile à l'échelle mondiale. (...) Selon Global Reach [une société promouvant la localisation des sites web], les groupes linguistiques se développant le plus vite sont les groupes non anglophones: on note une progression de 22,4% pour les sites web hispanophones, 12,3% pour les sites nippons, 14% pour les sites germanophones et 10% pour les sites francophones. On estime à 55,7 millions le nombre de personnes non anglophones ayant accès au web. (...) Alors que 6% seulement de la population mondiale est de langue maternelle anglaise (et 16% de langue maternelle espagnole), 80% des pages web sont en anglais.» Toujours selon Global Reach, 15% seulement des 500 millions d'habitants que compte l'Europe sont de langue maternelle anglaise, 28% maîtrisent bien l'anglais, et 32% consultent le web anglophone.

Le concept de «démocratie linguistique» est développé par Brian King, directeur du WorldWide Language Institute (WWLI), dans un entretien daté de septembre 1998: «Dans un rapport de l'UNESCO du début des années 1950, l'enseignement dispensé dans sa langue maternelle était considéré comme un droit fondamental de l'enfant. La possibilité de naviguer sur l'internet dans sa langue maternelle pourrait bien être son équivalent à l'Âge de l'Information. Si l'internet doit vraiment devenir le réseau mondial qu'on nous promet, tous les usagers devraient y avoir accès sans problème de langue. Considérer l'internet comme la chasse gardée de ceux qui, par accident historique, nécessité pratique ou privilège politique, connaissent l'anglais, est injuste à l'égard de ceux qui ne connaissent pas cette langue.»

Jean-Pierre Cloutier est l'auteur des *Chroniques de Cybérie*, une lettre d'information électronique sur l'actualité de l'internet. Il écrit en août 1999: «Cet été, le cap a été franchi. Plus de 50% des utilisateurs et utilisatrices du réseau sont hors des États-Unis. L'an prochain, plus de 50% des utilisateurs seront non anglophones. Il y a seulement cinq ans, c'était 5%. Formidable, non?»

Les usagers non anglophones atteignent en effet la barre des 50% au cours de l'été 2000. Selon Global Reach, ce pourcentage est de 52,5% en été 2001, 57% en décembre 2001, 59,8% en avril 2002, 64,4% en septembre 2003 (dont 34,9% d'Européens non anglophones et 29,4% d'Asiatiques) et 64,2% en mars 2004 (dont 37,9% d'Européens non anglophones et 33% d'Asiatiques).

De plus, nombre de communautés pratiquent le bilinguisme au quotidien, par exemple à Genève pour le français et l'allemand ou San Francisco pour l'anglais et l'espagnol, avec l'exigence de proposer toutes les informations en deux langues, le cas extrême étant la Communauté européenne avec ses 11 langues officielles en 2003, puis ses 24 langues officielles en 2007 après son élargissement progressif vers l'Europe de l'Est.

Vers un web multilingue

Premiers pas

Professeure d'espagnol en entreprise et traductrice, Maria Victoria Marinetti écrit en août 1999: «Il est très important de pouvoir communiquer en différentes langues. Je dirais même que c'est obligatoire, car l'information donnée sur l'internet est à destination du monde entier, alors pourquoi ne l'aurions-nous pas dans notre propre langue ou dans la langue que nous souhaitons utiliser? Information mondiale, mais pas de vaste choix dans les langues, ce serait contradictoire, pas vrai?»

Yahoo! est précurseur dans ce domaine. En Californie, deux étudiants de l'Université de Stanford, Jerry Lang et David Filo, lancent en janvier 1994 l'annuaire Yahoo! pour recenser les sites web et les classer par thèmes. L'annuaire est un succès, avec un classement plus pointu que celui de moteurs de recherche comme AltaVista, où ces tâches sont entièrement automatisées. Mais, si une recherche ne donne pas de résultat dans Yahoo!, elle est automatiquement aiguillée vers AltaVista, et réciproquement.

Trois ans plus tard, Yahoo! propose un classement en 63 grandes catégories thématiques, tout comme une interface en plusieurs langues: anglais, allemand, coréen, français, japonais, norvégien et suédois.

En décembre 1997, AltaVista est le premier moteur de recherche à lancer un service gratuit de traduction automatisée de l'anglais vers cinq autres langues (allemand, espagnol, français, italien et portugais) et vice versa, la page originale et la traduction apparaissant en vis-à-vis à l'écran. AltaVista Translation, surnommé aussi Babel Fish, est l'œuvre de SYSTRAN, une société franco-américaine pionnière dans le traitement automatique des langues. Babel Fish est alimenté par des dictionnaires multilingues comprenant 2,5 millions de termes.

Bien qu'ayant ses limites, avec une traduction de trois pages maximum et un texte traduit très approximatif, ce service est immédiatement très apprécié des douze millions d'usagers que compte l'internet de l'époque, dont un nombre croissant d'usagers non anglophones. Babel Fish ouvre aussi la voie à d'autres services du même genre - développés entre autres par Alis Technologies, Lernout & Hauspie, Globalink ou Softissimo - et contribue grandement au plurilinguisme du web.

La Kotoba Home Page

Yoshi Mikami, informaticien à Tokyo (Japon), lance en décembre 1995 le site bilingue anglais-japonais *The Languages of the World by Computers and the Internet*, plus connu sous le nom de Logos Home Page ou Kotoba Home Page. Le site donne un bref historique de chaque langue, ses caractéristiques, son système d'écriture, son jeu de caractères et enfin la configuration du clavier dans la langue donnée. Yoshi Mikami est

également co-auteur (avec Kenji Sekine et Nobutoshi Kohara) de *Pour un web multilingue*, publié en août 1997 en japonais par les éditions O'Reilly avant d'être traduit en anglais, en allemand et en français (version française parue en septembre 1998).

Yoshi écrit en décembre 1998: «Ma langue maternelle est le japonais. Comme j'ai suivi mes études de troisième cycle aux États-Unis et que j'ai travaillé dans l'informatique, je suis devenu bilingue japonais/anglais américain. J'ai toujours été intéressé par différentes langues et cultures, aussi j'ai appris le russe, le français et le chinois dans la foulée. A la fin de 1995, j'ai créé sur le web *The Languages of the World by Computers and the Internet* et j'ai tenté de donner - en anglais et en japonais - un bref historique de toutes ces langues, ainsi que les caractéristiques propres à chaque langue et à sa phonétique. Suite à l'expérience acquise, j'ai invité mes deux associés à écrire un livre sur la conception, la création et la présentation de pages web multilingues, livre qui fut publié en août 1997 dans son édition japonaise, le premier livre au monde sur un tel sujet.»

Comment voit-il l'évolution vers un web multilingue? «Il y a des milliers d'années de cela, en Égypte, en Chine et ailleurs, les gens étaient plus sensibles au fait de communiquer leurs lois et leurs réflexions non seulement dans une langue mais dans plusieurs. Dans notre monde moderne, chaque État a le plus souvent adopté une seule langue de communication. A mon avis, l'internet verra l'utilisation plus grande de langues différentes et de pages multilingues - et pas seulement une gravitation autour de l'anglais américain - et un usage plus créatif de la traduction informatique multilingue. 99% des sites web créés au Japon sont en japonais!»

OneLook Dictionaries

Robert Ware, enseignant, lance en avril 1996 le site OneLook Dictionaries pour permettre une recherche rapide dans des centaines de dictionnaires couvrant divers domaines: affaires, informatique et internet, médecine, religion, sciences et techniques, sports, généralités et argot.

Il explique en septembre 1998: «À titre personnel, je suis presque uniquement en contact avec des gens qui ne pratiquent qu'une langue et qui n'ont pas beaucoup de motivation pour développer leurs aptitudes linguistiques. Être en contact avec le monde entier change cette approche des choses. Et la change en mieux! (...) J'ai été long à inclure des dictionnaires non anglophones (en partie parce que je suis monolingue). Mais vous en trouverez maintenant quelques-uns.»

OneLook Dictionaries compte 2 millions de termes provenant de 425 dictionnaires en 1998, 2,5 millions de termes provenant de 530 dictionnaires en 2000, et 5 millions de termes provenant de 910 dictionnaires en 2003.

Robert Ware raconte aussi dans le même entretien: «Un fait intéressant s'est produit par le passé qui a été très instructif pour moi. En 1994, je travaillais pour un établissement scolaire et j'essayais d'installer un logiciel sur un modèle d'ordinateur particulier. J'ai trouvé une personne qui était en train de travailler sur le même problème, et nous avons commencé à échanger des courriels. Soudain, cela m'a frappé... Le logiciel avait été écrit à 40 kilomètres de là, mais c'était une personne située à l'autre bout du monde qui m'aidait. Les distances et l'éloignement

géographique n'importaient plus! Et bien, ceci est formidable, mais à quoi cela nous mène-t-il? Je ne puis communiquer qu'en anglais mais, heureusement, mon correspondant pouvait utiliser aussi bien l'anglais que l'allemand qui était sa langue maternelle. L'internet a supprimé une barrière, celle de la distance, mais il subsiste la barrière de la langue, bien réelle.

Il semble que l'internet propulse simultanément les gens dans deux directions différentes. L'internet, anglophone à l'origine, relie les gens dans le monde entier. Par là même, il favorise une langue commune pour communiquer. Mais il suscite aussi des contacts entre des personnes de langue différente et permet ainsi de développer un intérêt plus grand pour le multilinguisme. Si une langue commune est appréciable, elle ne remplace en aucun cas la nécessité de plusieurs langues. L'internet favorise ainsi à la fois une langue commune et le multilinguisme, et ceci est un facteur qui aide à trouver des solutions. L'intérêt croissant pour les langues et le besoin qu'on en a stimulent de par le monde la création de cours de langues et d'instruments d'aide linguistique, et l'internet fournit la possibilité de les rendre disponibles rapidement et à bon marché.»

Autres expériences

Même si l'anglais est encore prédominant à la fin des années 1990, les sites bilingues ou plurilingues sont de plus en plus nombreux, ce pour des raisons aussi bien commerciales que culturelles, qui prennent en compte le fait que tout le monde ne comprend pas l'anglais.

Brian King, directeur du WorldWide Language Institute (WWLI), écrit en septembre 1998: «De même que l'utilisateur non anglophone peut maintenant avoir accès aux nouvelles technologies dans sa propre langue, l'impact du commerce électronique peut constituer une force majeure qui fasse du multilinguisme la voie la plus naturelle vers le cyberespace. Les vendeurs de produits et services dans le marché virtuel mondial que devient l'internet doivent être préparés à desservir un monde virtuel qui soit aussi multilingue que le monde physique. S'ils veulent réussir, ils doivent s'assurer qu'ils parlent bien la langue de leurs clients!»

Le réseau ELSNET (European Network in Language and Speech) regroupe une centaine de partenaires académiques et commerciaux, l'objectif étant de mettre sur pied des systèmes multilingues pour la langue parlée et la langue écrite.

Steven Krauwer, coordinateur d'ELSNET, explique en septembre 1998: «En tant que citoyen européen, je pense que le multilinguisme sur le web est absolument essentiel. À mon avis, ce n'est pas une situation saine à long terme que seuls ceux qui ont une bonne maîtrise de l'anglais puissent pleinement exploiter les bénéfices du web. En tant que chercheur (spécialisé dans la traduction automatique), je vois le multilinguisme comme un défi majeur: pouvoir garantir que l'information sur le web soit accessible à tous, indépendamment des différences de langue. (...) Je compte passer le reste de ma vie professionnelle à utiliser les technologies de l'information pour supprimer ou au moins réduire la barrière des langues.»

Il ajoute en août 1999: «Je suis de plus en plus convaincu que nous devons veiller à ne pas aborder le problème du multilinguisme en l'isolant du reste. Je reviens de France, où

j'ai passé de très bonnes vacances d'été. Même si ma connaissance du français est sommaire (c'est le moins que l'on puisse dire), il est surprenant de voir que je peux malgré tout communiquer sans problème en combinant ce français sommaire avec des gestes, des expressions du visage, des indices visuels, des schémas, etc. Je pense que le web (contrairement au système vieillot du courriel textuel) peut permettre de combiner avec succès la transmission des informations par différents canaux (ou différents moyens), même si ce processus n'est que partiellement satisfaisant pour chacun des canaux pris isolément.»

Pour un véritable multilinguisme sur le web, Steven Krauwer suggère plusieurs solutions pratiques: «(a) en ce qui concerne les auteurs: une meilleure formation des auteurs de sites web pour exploiter les combinaisons possibles permettant d'améliorer la communication en surmontant la barrière de la langue (et pas seulement par un vernis superficiel); (b) en ce qui concerne les usagers: des logiciels de traduction de type AltaVista Translation, dont la qualité n'est pas frappante, mais qui a le mérite d'exister; (c) en ce qui concerne les logiciels de navigation: des logiciels de traduction intégrée, particulièrement pour les langues non dominantes, et des dictionnaires intégrés plus rapides à consulter.»

L'affaire de tous

Le multilinguisme est l'affaire de tous, témoin cet Appel du Comité européen pour le respect des cultures et des langues en Europe (CERCLE) qui, diffusé en 1998 dans les onze langues officielles de l'Union européenne (allemand, anglais, danois, espagnol, finlandais, français, grec, hollandais, italien, portugais et suédois), défend «une Europe humaniste, plurilingue et riche de sa diversité culturelle». Le CERCLE propose aux réviseurs du Traité de l'Union européenne douze amendements prenant en compte le respect des cultures et des langues.

On lit dans cet Appel que «la diversité et le pluralisme linguistiques ne sont pas un obstacle à la circulation des hommes, des idées et des marchandises ou services, comme veulent le faire croire certains, alliés objectifs, conscients ou non, de la culture et de la langue dominantes. C'est l'uniformisation et l'hégémonie qui sont un obstacle au libre épanouissement des individus, des sociétés et de l'économie de l'immatériel, source principale des emplois de demain. Le respect des langues, à l'inverse, est la dernière chance pour l'Europe de se rapprocher des citoyens, objectif toujours affiché, presque jamais mis en pratique. L'Union doit donc renoncer à privilégier la langue d'un seul groupe.»

Bruno Didier, webmestre de la bibliothèque de l’Institut Pasteur, écrit en août 1999: «L'internet n’est une propriété ni nationale, ni linguistique. C’est un vecteur de culture, et le premier support de la culture, c’est la langue. Plus il y a de langues représentées dans leur diversité, plus il y aura de cultures sur l'internet. Je ne pense pas qu’il faille justement céder à la tentation systématique de traduire ses pages dans une langue plus ou moins universelle. Les échanges culturels passent par la volonté de se mettre à la portée de celui vers qui on souhaite aller. Et cet effort passe par l’appréhension de sa langue. Bien entendu c’est très utopique comme propos. Concrètement, lorsque je fais de la veille, je peste dès que je rencontre des sites norvégiens ou brésiliens sans un minimum d’anglais.»

En 2000, le web est multilingue, mais la barrière de la langue est loin d'avoir disparu. Si toutes les langues sont désormais représentées sur le web, on oublie trop souvent que de nombreux usagers sont unilingues, et que même les polyglottes ne peuvent connaître toutes les langues. Il importe aussi d'avoir à l'esprit l'ensemble des langues, et pas seulement les langues dominantes. Il reste à créer des passerelles entre les communautés linguistiques pour favoriser la circulation des écrits dans d'autres langues, notamment en améliorant la qualité des logiciels de traduction.

L'anglais reste prédominant

Une réalité statistique
Commentaires francophones
Des solutions logicielles

Une réalité statistique

L'anglais reste prépondérant et ceci n'est pas près de disparaître. Comme indiqué en janvier 1999 par Marcel Grangier, responsable de la section française des services linguistiques centraux de l'Administration fédérale suisse, «cette suprématie n'est pas un mal en soi, dans la mesure où elle résulte de réalités essentiellement statistiques (plus de PC par habitant, plus de locuteurs de cette langue, etc.). La riposte n'est pas de "lutter contre l'anglais" et encore moins de s'en tenir à des jérémiades, mais de multiplier les sites en d'autres langues. Notons qu'en qualité de service de traduction, nous préconisons également le multilinguisme des sites eux-mêmes. La multiplication des langues présentes sur l'internet est inévitable, et ne peut que bénéficier aux échanges multiculturels.»

Professeur en technologies de la communication à la Webster University de Genève (Suisse), Henk Slettenhaar insiste tout autant sur la nécessité de sites bilingues, dans la langue originale et en anglais. «Les communautés locales présentes sur le web devraient en tout premier lieu utiliser leur langue pour diffuser des informations, écrit-il en décembre 1998. Si elles veulent présenter ces informations à la communauté mondiale, celles-ci doivent être disponibles aussi en anglais. Je pense qu'il existe un réel besoin de sites bilingues. (...) Mais je suis enchanté qu'il existe maintenant tant de documents disponibles dans leur langue originale. Je préfère de beaucoup lire l'original avec difficulté plutôt qu'une traduction médiocre.»

Henk ajoute en août 1999: «A mon avis, il existe deux types de recherches sur le web. La première est la recherche globale dans le domaine des affaires et de l'information. Pour cela, la langue est d'abord l'anglais, avec des versions locales si nécessaire. La seconde, ce sont les informations locales de tous ordres dans les endroits les plus reculés. Si l'information est à destination d'une ethnie ou d'un groupe linguistique, elle doit d'abord être disponible dans la langue de l'ethnie ou du groupe, avec peut-être un résumé en anglais.»

Guy Antoine, créateur de Windows on Haiti, un site de référence sur la culture haïtienne, croit en la nécessité de l'anglais en tant que langue commune. «Pour des raisons pratiques, l'anglais continuera de dominer le web, relate-t-il en novembre 1999. Je ne pense pas que ce soit une mauvaise chose, en dépit des sentiments régionalistes qui s'y opposent, parce que nous avons besoin d'une langue commune permettant de favoriser les communications à l'échelon international. Ceci dit, je ne partage pas l'idée pessimiste selon laquelle les autres langues n'ont plus qu'à se soumettre à la langue dominante. Au contraire. Tout d'abord l'internet peut héberger des informations utiles sur les langues minoritaires, qui seraient autrement amenées à disparaître sans laisser de traces. De plus, à mon avis, l'internet incite les gens à apprendre les langues associées aux cultures qui les intéressent. Ces personnes réalisent rapidement que la langue d'un peuple est un élément fondamental de sa culture. De ce fait, je n'ai pas

grande confiance dans les outils de traduction automatique qui, s'ils traduisent les mots et les expressions, ne peuvent guère traduire l'âme d'un peuple. Que sont les Haïtiens, par exemple, sans le kreyòl (créole pour les non initiés), une langue qui a permis de souder entre elles diverses tribus africaines transplantées à Haïti pendant la période de l'esclavage? Cette langue représente de manière la plus palpable l'unité de notre peuple. Elle est toutefois principalement une langue parlée et non écrite. A mon avis, le web va changer cet état de fait plus qu'aucun autre moyen traditionnel de diffusion d'une langue. Dans Windows on Haiti, la langue principale est l'anglais, mais on y trouve tout aussi bien un forum de discussion animé conduit en kreyòl. On y trouve également des documents sur Haïti en français et dans l'ancien créole colonial, et je suis prêt à publier d'autres documents en espagnol et dans diverses langues. Je ne propose pas de traductions, mais le multilinguisme est effectif sur ce site, et je pense qu'il deviendra de plus en plus la norme sur le web.»

Geoffrey Kingscott est directeur général de Praetorius, une société britannique spécialisée en linguistique appliquée. Il explique en septembre 1998: «Les caractéristiques propres au web sont la multiplicité de générateurs de sites et le bas prix de l'émission de messages. Ceci favorisera donc le multilinguisme au fur et à mesure du développement du web. Comme celui-ci a vu le jour aux États-Unis, il est encore principalement en anglais, mais ce n'est qu'un phénomène temporaire. Pour expliquer ceci plus en détail, je dirais que, quand nous comptions sur l'imprimé ou l'audiovisuel (film, télévision, radio, vidéo, cassettes), l'information ou le divertissement que nous attendions dépendait d'agents (éditeurs, stations de télévision ou de radio, producteurs de cassettes ou de vidéos) qui devaient subsister commercialement et, dans le cas de la radio-télédiffusion du service public, avec de sévères contraintes budgétaires. Ceci signifie que la quantité de clients est primordiale, et détermine la nécessité de langues autres que l'omniprésent anglais. Ces contraintes disparaissent avec le web.

Pour ne donner qu'un exemple mineur tiré de notre expérience, nous publions la version imprimée de notre magazine *Language Today* uniquement en anglais, qui est le dénominateur commun de nos lecteurs. Quand nous utilisons un article qui était originellement dans une autre langue que l'anglais, ou que nous relatons un entretien mené dans une autre langue que l'anglais, nous le traduisons en anglais et nous ne publions que la version anglaise, pour la raison suivante: le nombre de pages que nous pouvons imprimer est limité, et déterminé en fonction de notre clientèle (annonceurs et abonnés). Par contre, dans notre version web, nous proposons aussi la version originale.»

Commentaires francophones

Luc dall'Armellina, co-auteur et webmestre d’oVosite, un espace d’écriture hypermédia, écrit en juin 2000: «L'anglais s'impose sans doute parce qu'il est devenu la langue commerciale d'échange généralisée; il semble important que toutes les langues puissent continuer à être représentées parce que chacune d'elle est porteuse d'une vision "singulière" du monde. La traduction simultanée (proposée par AltaVista par exemple) ou les versions multilingues d'un même contenu me semblent aujourd'hui les meilleures réponses au danger de pensée unique que représenterait une seule langue d'échange. Peut-être appartient-il aux éditeurs des systèmes d'exploitation (ou de navigateurs?) de proposer des solutions de traduction partielle, avec toutes les limites

connues des systèmes automatiques de traduction...»

Pierre François Gagnon, fondateur d'Éditel et pionnier de l'édition littéraire francophone en ligne, écrit en juillet 2000: «Je pense que, si les diverses langues de la planète vont occuper chacune l'internet en proportion de leur poids démographique respectif, la nécessité d'une langue véhiculaire unique se fera sentir comme jamais auparavant, ce qui ne fera qu'assurer davantage encore la suprématie planétaire de l'anglais, ne serait-ce que du fait qu'il a été adopté définitivement par l'Inde et la Chine. Or la marche de l'histoire n'est pas plus comprimable dans le dé à coudre d'une quelconque équation mathématique que le marché des options en bourse!»

Philippe Loubière, traducteur littéraire et dramatique, dénonce pour sa part la main-mise anglophone sur le réseau. «Tout ce qui peut contribuer à la diversité linguistique, sur internet comme ailleurs, est indispensable à la survie de la liberté de penser, explique-t-il en mars 2001. Je n'exagère absolument pas: l'homme moderne joue là sa survie. Cela dit, je suis très pessimiste devant cette évolution. Les Anglo-saxons vous écrivent en anglais sans vergogne. L'immense majorité des Français constate avec une indifférence totale le remplacement progressif de leur langue par le mauvais anglais des marchands et des publicitaires, et le reste du monde a parfaitement admis l'hégémonie linguistique des Anglo-saxons parce qu'ils n'ont pas d'autres horizons que de servir ces riches et puissants maîtres. La seule solution consisterait à recourir à des législations internationales assez contraignantes pour obliger les gouvernements nationaux à respecter et à faire respecter la langue nationale dans leur propre pays (le français en France, le roumain en Roumanie, etc.), cela dans tous les domaines et pas seulement sur internet. Mais ne rêvons pas...»

C'est aussi le sentiment de Blaise Rosnay, webmestre du site du Club des poètes, qui écrit en janvier 2000: «Dans la mesure où la culture française, y compris contemporaine, pourra être diffusée sans obstacles, la langue française aura la possibilité de rester vivante sur le réseau. Ses œuvres, liées au génie de notre langue, susciteront nécessairement de l'intérêt puisqu'elles sont en prise avec l'évolution actuelle de l'esprit humain. Dans la mesure où il y aura une volonté d'utiliser l'internet comme moyen de partage de la connaissance, de la beauté, de la culture, toutes les langues, chacune avec leur génie propre, y auront leur place. Mais si l'internet, comme cela semble être le cas, abandonne ces promesses pour devenir un lieu unique de transactions commerciales, la seule langue qui y sera finalement parlée sera une sorte de jargon dénaturant la belle langue anglaise, je veux dire un anglais amoindri à l'usage des relations uniquement commerciales.»

Richard Chotin, professeur à l'École supérieure des affaires (ESA) de Lille, rappelle à juste titre que la suprématie de l'anglais a succédé à celle du français. «Le problème est politique et idéologique, explique-t-il en septembre 2000. C'est celui de l'"impérialisme" de la langue anglaise découlant de l'impérialisme américain. Il suffit d'ailleurs de se souvenir de l'"impérialisme" du français aux 18e et 19e siècles pour comprendre la déficience en langues des étudiants français: quand on n'a pas besoin de faire des efforts pour se faire comprendre, on n'en fait pas, ce sont les autres qui les font.»

Bakayoko Bourahima, bibliothécaire de l'École nationale supérieure de statistique et d'économie appliquée (ENSEA) d'Abidjan (Côte d'Ivoire), écrit en juillet 2000: «Pour nous les Africains francophones, le diktat de l'anglais sur la toile représente pour la masse un

double handicap d’accès aux ressources du réseau. Il y a d’abord le problème de l’alphabétisation qui est loin d’être résolu et que l’internet va poser avec beaucoup plus d’acuité, ensuite se pose le problème de la maîtrise d’une seconde langue étrangère et son adéquation à l’environnement culturel. En somme, à défaut de multilinguisme, l’internet va nous imposer une seconde colonisation linguistique avec toutes les contraintes que cela suppose. Ce qui n’est pas rien quand on sait que nos systèmes éducatifs ont déjà beaucoup de mal à optimiser leurs performances, en raison, selon certains spécialistes, des contraintes de l’utilisation du français comme langue de formation de base. Il est donc de plus en plus question de recourir aux langues vernaculaires pour les formations de base, pour "désenclaver" l’école en Afrique et l’impliquer au mieux dans la valorisation des ressources humaines. Comment faire? Je pense qu’il n’y a pas de chance pour nous de faire prévaloir une quelconque exception culturelle sur la toile, ce qui serait de nature tout à fait grégaire. Il faut donc que les différents blocs linguistiques s’investissent beaucoup plus dans la promotion de leur accès à la toile, sans oublier leurs différentes spécificités internes.»

Lucie de Boutiny, romancière, écrit en septembre 2000: "Les chiffres de septembre 2000 montrent que 51% des utilisateurs sont anglo-saxons, et 78% des sites aussi. Les chiffres de cette prépondérance baissent à mesure qu'augmentent le nombre des internautes de par le monde... L'anglais va devenir la deuxième langue mondiale après la langue natale, mais il y en aura d'autres. Un exemple: personnellement, à l'âge de 4 ans, je parlais trois langues alors que je ne savais ni lire ni écrire. Pour parler une langue, il peut suffire d'avoir la chance de l'écouter. On peut espérer que le cosmopolitisme traverse toutes les classes sociales en raison, par exemple, de l'Union européenne, du nomadisme des travailleurs, de la facilité de déplacement à l'étranger des étudiants, de la présence des chaînes TV et sites étrangers, etc."

Gérard Fourestier, créateur de Rubriques à Bac, un site destiné aux étudiants du premier cycle universitaire, écrit en octobre 2000: «Je suis de langue française. J'ai appris l'allemand, l'anglais, l'arabe, mais je suis encore loin du compte quand je surfe dans tous les coins de la planète. Il serait dommage que les plus nombreux ou les plus puissants soient les seuls qui "s'affichent" et, pour ce qui est des logiciels de traduction, il y a encore largement à faire. (...) Pour l'instant, [il importe] de connaître suffisamment d'anglais et de créer beaucoup plus encore en français.»

Des solutions logicielles

Alain Bron, consultant en systèmes d'information et écrivain, explique pour sa part en novembre 1999: «Il y aura encore pendant longtemps l'usage de langues différentes et tant mieux pour le droit à la différence. Le risque est bien entendu l'envahissement d'une langue au détriment des autres, donc l'aplanissement culturel. Je pense que des services en ligne vont petit à petit se créer pour pallier cette difficulté. Tout d'abord, des traducteurs pourront traduire et commenter des textes à la demande, et surtout les sites de grande fréquentation vont investir dans des versions en langues différentes, comme le fait l'industrie audiovisuelle.»

Tôt ou tard, le pourcentage des langues sur le réseau correspondra-t-il à leur répartition sur la planète? Rien n’est moins sûr à l’heure de la fracture numérique entre riches et pauvres, entre zones rurales et zones urbaines, entre régions favorisées et régions défavorisées, entre l’hémisphère nord et l’hémisphère sud, entre pays développés et

pays en développement.

Selon Zina Tucsnak, ingénieur d’études au laboratoire ATILF (Analyse et traitement informatique de la langue française), interviewée en octobre 2000, "le meilleur moyen serait l’application d’une loi par laquelle on va attribuer un "quota" à chaque langue. Mais n’est-ce pas une utopie de demander l’application d’une telle loi dans une société de consommation comme la nôtre?"

À la même date, Emmanuel Barthe, documentaliste juridique, exprime un avis contraire: «Des signes récents laissent penser qu’il suffit de laisser les langues telles qu’elles sont actuellement sur le web. En effet, les langues autres que l’anglais se développent avec l’accroissement du nombre de sites web nationaux s’adressant spécifiquement aux publics nationaux, afin de les attirer vers internet. Il suffit de regarder l’accroissement du nombre de langues disponibles dans les interfaces des moteurs de recherche généralistes. Il serait néanmoins utile (et bénéfique pour un meilleur équilibre des langues) de disposer de logiciels de traduction automatique de meilleure qualité et à très bas prix sur internet. La récente mise sur le web du GDT (Grand dictionnaire terminologique, rédigé par l'Office de la langue française du Québec) va dans ce sens.»

Michel Benoît, romancier vivant à Montréal (Québec), écrit en juin 2000: «Lorsqu'un problème affecte une structure, quelle qu'elle soit, j'ai toujours tendance à imaginer que c'est techniquement que le problème trouve sa solution. Vous connaissez cette théorie? Si les Romains avaient trouvé le moyen d'enlever le plomb de leur couvert d'étain, Néron ne serait jamais devenu fou et n'aurait jamais incendié Rome. Escusi, farfelu? Peut-être que oui, peut-être que non. E que save? L'internet multilingue? Demain, ou après-demain au plus. Voyons, pensez au premier ordinateur, il y a de cela un peu plus que cinquante ans. Un étage au complet pour faire à peine plus que les quatre opérations de base. Dans ce temps-là, un bug, c'était véritablement une mouche - ou autre insecte - qui s'insérait entre les lecteurs optiques. De nos jours [en 2000], un carte de 3 cm x 5 cm fait la même chose. La traduction instantanée: demain, après-demain au plus.»

C'est aussi l'avis de Pierre Magnenat, responsable de la cellule «gestion et prospective» du centre informatique de l'Université de Lausanne, qui écrit en octobre 2000: «La seule solution que je vois serait qu'un effort majeur et global soit entrepris pour développer des traducteurs automatiques. Je ne pense pas qu'une quelconque incitation ou autre quota pourrait empêcher la domination totale de l'anglais. Cet effort pourrait - et devrait - être initié au niveau des États, et disposer des moyens suffisants pour aboutir.»

Pierre-Noël Favennec, expert à la direction scientifique de France Télécom R&D, souligne en février 2001: «Les recherches sur la traduction automatique devraient permettre une traduction automatique dans les langues souhaitées, mais avec des applications pour toutes les langues et non les seules dominantes (ex.: diffusion de documents en japonais, si l’émetteur est de langue japonaise, et lecture en breton, si le récepteur est de langue bretonne...). Il y a donc beaucoup de travaux à faire dans le domaine de la traduction automatique et écrite de toutes les langues.»

Le français sur le web

Une «nation de langue» francophone
Le patrimoine linguistique en ligne

Une «nation de langue» francophone

En décembre 1997, Tim Berners-Lee, l'inventeur du World Wide Web (en 1990), déclare à Pierre Ruetschi, journaliste à la *Tribune de Genève*, un quotidien suisse: «Pourquoi les Francophones ne mettent-ils pas davantage d'informations sur le web? Est-ce qu'ils pensent que personne ne veut la lire, que la culture française n'a rien à offrir? C'est de la folie, l'offre est évidemment énorme.»

C'est chose faite dans les années qui suivent, d'abord par le Québec, une région dynamique qui dispose d'un accès beaucoup plus facile et économique à l'internet les premières années, et ensuite par la Francophonie dans son ensemble.

Début 1998, les Québécois, pionniers de l'internet francophone, attendent de pied ferme l'arrivée en masse de sites web français, y compris commerciaux. Lors d'un entretien publié par le magazine en ligne *Multimédium*, Louise Beaudouin, ministre de la Culture et des Communications au Québec, déclare en février 1998: «J'attendais depuis deux ans que la France se réveille. Aujourd'hui, je ne m'en plaindrai pas.» À cette date, le Québec (avec 6 millions d'habitants) propose plus de sites web que la France (avec 60 millions d'habitants). La ministre attribue le retard de la France à deux facteurs: d'une part les tarifs élevés du téléphone (et donc de l'internet, puisque la connexion s'effectue par le biais de la ligne téléphonique), d'autre part les transactions commerciales possibles sur le minitel (le videotex français) depuis plusieurs années, ce qui ralentit l'expansion du commerce électronique sur l'internet.

C'est l'UREC (Unité réseaux du Centre national de la recherche scientifique) qui, en France, lance le premier annuaire de sites web francophones. L'annuaire de l'UREC a pour but de se familiariser avec le web sans se noyer dans la masse d'informations mondiale, et de connaître les sites qui petit à petit fleurissent en langue française. Créé début 1994, il recense d'abord les sites académiques avant de devenir plus généraliste. D'autres annuaires voient ensuite le jour, dont certains débutés avec l'aide de l'UREC. Le nombre de sites web, y compris commerciaux, augmente de manière exponentielle, si bien que la gestion d'un annuaire généraliste devient difficile. En juillet 1997, considérant sa mission comme accomplie, l'UREC arrête la mise à jour de cet annuaire généraliste, et le remplace par un annuaire spécialisé consacré à l'enseignement supérieur et à la recherche.

Fondée en mars 1970 à Niamey (Niger) pour regrouper 21 États francophones, l'Agence la Francophonie - qui se prénomme d'abord l'Agence de coopération culturelle et technique (ACCT) - compte 47 États francophones en 1997. Cette agence se veut un «instrument de coopération multilatérale née d'un idéal, celui de créer une communauté qui fasse entendre sa voix dans le concert des nations».

Une Conférence des ministres francophones chargés des inforoutes a lieu à Montréal en mai 1997. Datée du 21 mai 1997, la Déclaration de Montréal propose de «développer

une aire francophone d'éducation, de formation et de recherche; soutenir la création et la circulation de contenus francophones et contribuer à la sauvegarde et à la valorisation des patrimoines; encourager la promotion de l'aire francophone de développement économique; mettre en place une vigie francophone (veille active); sensibiliser prioritairement la jeunesse ainsi que les utilisateurs, les producteurs et les décideurs; assurer la présence et la concertation des Francophones dans les instances spécialisées.»

Par ailleurs, l'Agence universitaire de la Francophonie (AUF) crée le réseau internet REFER pour desservir la communauté scientifique et technique en Afrique, en Asie et en Europe orientale, avec 24 pays participants en 2002.

Le 20 mars 2010, la Francophonie fête ses quarante ans d'existence. L'Organisation internationale de la Francophonie (OIF) rassemble désormais 70 États et gouvernements (56 membres et 14 observateurs) totalisant 870 millions d'habitants, soit 13% de la population mondiale.

Mais, selon diverses sources, on ne recenserait en fait que 200 millions de Francophones dans le monde, dont 72 millions de «locuteurs partiels». 96,2 millions de Francophones vivraient en Afrique. Le français serait la neuvième langue de la planète et 5% des pages internet seraient rédigées en français. Tous chiffres à prendre avec précaution.

S'il est la langue des pays francophones, le français est aussi la deuxième langue utilisée dans les organisations internationales. Malgré la pression anglophone - réelle ou supposée selon les cas -, des Francophones veillent à ce que leur langue ait une place significative dans le monde, au même titre que les autres grandes langues de communication que sont l'anglais, l'arabe, le chinois et l'espagnol. Là aussi, l'optique est aussi bien la défense d'une langue que le respect du multilinguisme et de la diversité des peuples.

Le patrimoine linguistique en ligne

L'internet peut favoriser non seulement l'usage du français, en cas de volonté politique et culturelle dans ce sens, mais aussi la transmission du patrimoine linguistique. Un exemple parmi d'autres est celui des dictionnaires anciens, qui trouvent une nouvelle vie grâce au web.

Dès 1996, le site de l'Institut national de la langue française (InaLF) offre des ressources terminologiques sur le discours littéraire des 14e au 20e siècles (contenu, sémantique, thématique), sur la langue courante (langue écrite, langue parlée, argot) et sur le discours scientifique et technique.

Christiane Jadelot, ingénieur d'études à l'INaLF-Nancy, explique en juin 1998: «Les premières pages sur l'INaLF ont été mises sur l'internet au milieu de l'année 1996, à la demande de Robert Martin, directeur de l'INaLF. J'ai participé à la mise sous internet de ces pages (...). La direction a senti la nécessité urgente de nous faire connaître par l'internet, que beaucoup d'autres entreprises utilisaient déjà pour promouvoir leurs produits. Nous sommes en effet "Unité de recherche et de service" et nous avons donc à trouver des clients pour nos produits informatisés, le plus connu d'entre eux étant la

base textuelle FRANTEXT [sur l'internet depuis début 1995, NDLR], ainsi qu'une maquette du tome 14 du TLF [Trésor de la langue française]. Il était donc nécessaire de faire connaître l'ensemble de l'INaLF par ce moyen. Cela correspondait à une demande générale.»

La base FRANTEXT comprend, en mode interactif, 180 millions de mots-occurrences provenant d'une collection représentative de 3.500 unités textuelles en arts, sciences et techniques des 16e-20e siècles. Début 1998, 82 centres de recherche et bibliothèques universitaires sont abonnés, en Europe, en Australie, au Japon et au Canada, ce qui représente 1.250 postes de travail ayant accès à la base, avec une cinquantaine de sessions d'interrogations par jour.

Le Projet ARTFL (ARTFL: American and French Research on the Treasury of the French Language) est un projet commun du Centre national de la recherche scientifique (CNRS, France) et de l'Université de Chicago (Illinois, États-Unis). À l'origine, ce projet a pour but de constituer une base de données de 2.000 textes des 13e-20e siècles ayant trait à la littérature, la philosophie, les arts et les sciences.

En 1998, l'ARTFL travaille à la version en ligne exhaustive de la première édition (1751-1772) de l'*Encyclopédie ou Dictionnaire raisonné des sciences, des métiers et des arts* de Diderot et d'Alembert. 72.000 articles rédigés par plus de 140 collaborateurs - dont Voltaire, Rousseau, d'Alembert, Marmontel, d'Holbach, Turgot, etc. - ont fait de cette encyclopédie un monumental ouvrage de référence pour les arts et les sciences. Destinée à rassembler puis divulguer les connaissances de l'époque, elle porte la marque des courants intellectuels et sociaux du 18e siècle, et c'est grâce à elle qu'ont été propagées les idées du Siècle des Lumières.

L'*Encyclopédie* comprend 17 volumes de texte - qui représentent 18.000 pages et 20.736.912 mots - et 11 volumes de planches.

La base de données correspondant au premier volume est accessible en ligne à titre expérimental. La recherche peut être effectuée par mot, portion de texte, auteur ou catégorie, ou par la combinaison de ces critères entre eux. On dispose de renvois d'un article à l'autre, au moyen de liens permettant d'aller d'une planche au texte ou du texte au fac-similé des pages originales. L'automatisation complète des procédures de saisie entraîne des erreurs typographiques et des erreurs d'identification qui sont ensuite corrigées au fil des mois. La recherche d'images par mot, portion de texte ou catégorie est également possible dans un deuxième temps.

L'ARTFL travaille aussi à un projet de base de données pour le *Dictionnaire de l'Académie française*, dont les différentes éditions se sont échelonnées entre 1694 et 1935. Ce projet inclut la saisie et l'édition du texte, ainsi que la création d'un moteur de recherche spécifique. La première édition (1694) et la cinquième édition (1798) du dictionnaire sont les premières à être disponibles pour une recherche par mot, puis pour une recherche en texte intégral. Les différentes éditions sont ensuite combinées dans une base de données unique, qui permet de juger de l'évolution d'un terme en consultant aussi bien une édition particulière que l'ensemble des éditions.

Les autres projets de l'ARTFL sont la version image de l'édition de 1740 du *Dictionnaire historique et critique* de Philippe Bayle, le *Roget's Thesaurus* de 1911, le *Webster's*

Revised Unabridged Dictionary de 1913, le *Thresor de la langue française* de Jean Nicot (1606), un projet multilingue sur La Bible comprenant entre autres *La Bible française* de Louis Segond (1910), etc.

Premiers projets multilingues

Travlang
Les projets de Tyler
NetGlos
Logos

Travlang

Travlang, site dédié à la fois aux voyages et aux langues, est créé par Michael C. Martin en 1994 sur le site de son université alors qu'il était étudiant en physique. Devenu chercheur au Lawrence Berkeley National Laboratory (Californie), Michael Martin poursuit la gestion de ce site devenu très populaire.

La section Foreign Languages for Travelers donne la possibilité d'apprendre les rudiments de soixante langues sur le web. La section Translating Dictionaries donne accès à des dictionnaires gratuits dans diverses langues (afrikaans, allemand, danois, espagnol, espéranto, finnois, français, frison, hollandais, hongrois, italien, latin, norvégien, portugais et tchèque). Ces dictionnaires sont le plus souvent sommaires et de qualité inégale. Le site offre aussi de nombreux liens vers des services de traduction, des écoles de langue, des librairies multilingues, etc.

Michael Martin écrit en août 1998: «Je pense que le web est un endroit idéal pour rapprocher les cultures et les personnes, et ceci inclut d'être multilingue. Notre site Travlang est très populaire pour cette raison, et les gens aiment le contact avec d'autres parties du monde. (...) L'internet est vraiment un outil important pour communiquer avec des gens avec lesquels on n'aurait pas l'occasion de dialoguer autrement. J'apprécie vraiment la collaboration générale qui a rendu possibles les pages de Foreign Languages for Travelers. (...) Je pense que les traductions intégrales informatisées vont devenir monnaie courante, et qu'elles permettront de communiquer à la base avec davantage de gens. Ceci aidera aussi à amener davantage l'internet au monde non anglophone.»

Les projets de Tyler

Créée par Tyler Chambers en mai 1994, la Human-Languages Page est un catalogue détaillé de 1.800 ressources linguistiques dans une centaine de langues. Ces ressources sont classées dans plusieurs rubriques: langues et littérature, écoles et institutions, ressources linguistiques, produits et services, organismes, emplois et stages, dictionnaires et cours de langues.

Tyler Chambers mène aussi un autre projet relatif aux langues, l'Internet Dictionary Project, un projet coopératif ouvert à tous pour la constitution de dictionnaires en accès libre sur le web, de l'anglais vers d'autres langues (allemand, espagnol, français, italien, latin et portugais).

Comme expliqué sur le site web, «le but de l'Internet Dictionary Project est de créer des dictionnaires de traduction grâce à l'aide des internautes. Ce site permet aux usagers du monde entier de consulter et de participer à la traduction de termes anglais dans d'autres langues. Les listes de termes anglais et leurs correspondants dans d'autres

langues sont ensuite mis à la disposition de tous sur ce site, sans restriction d'aucune sorte. (...) L'Internet Dictionary Project a débuté en 1995 pour combler une lacune et procurer des dictionnaires de traduction gratuits à la communauté des internautes et à tous ceux qui s'intéressent à l'informatique. Non seulement il est très utile d'avoir immédiatement accès à des dictionnaires par le World Wide Web, mais ceci permet aussi le développement de logiciels pouvant tirer parti de tels dictionnaires, que ce soit des programmes de traduction ou des vérificateurs d'orthographe ou encore des guides d'apprentissage des langues. En facilitant la création de ces dictionnaires en ligne par des milliers de volontaires, et en les mettant gratuitement à la disposition de tous, l'Internet Dictionary Project espère imprimer sa marque sur l'internet et susciter d'autres projets qui seront plus bénéfiques que de générer des revenus purement financiers.»

Tyler écrit en septembre 1998 lors d'un entretien par courriel: «Le multilinguisme sur le web était inévitable bien avant que ce médium ne se développe vraiment. Mon premier vrai contact avec l'internet date de 1994, un peu après ses débuts mais bien avant son expansion. 1994 a été aussi l'année où j'ai débuté mon premier projet web multilingue [The Human-Languages Page, NDLR], et il existait déjà un nombre significatif de ressources linguistiques en ligne. Ceci était antérieur à la création de Netscape. Mosaic était le seul navigateur sur le web, et les pages web étaient essentiellement des documents textuels reliés par des hyperliens. Avec l'amélioration des navigateurs et l'expérience acquise par les usagers, je ne pense pas qu'il existe une langue vivante qui ne soit pas maintenant représentée sur le web, que ce soit la langue des Indiens d'Amérique ou les dialectes moyen-orientaux. De même une pléthore de langues mortes peut maintenant trouver une audience nouvelle avec des érudits et autres spécialistes en ligne. À ma connaissance, très peu de jeux de caractères ne sont pas disponibles en ligne: les navigateurs ont maintenant la possibilité de visualiser les caractères romains, asiatiques, cyrilliques, grecs, turcs, etc. Accent Software a un produit appelé "Internet avec accents" qui serait capable de visualiser plus de 30 encodages différents. S'il existe encore des obstacles à la diffusion d'une langue spécifique sur le web, ceci ne devrait pas durer.»

En ce qui concerne les projets en ligne de Tyler: «Mon activité en ligne a été de rendre l'information linguistique accessible à davantage de gens par le biais de deux de mes projets sur le web [Human-Languages Page et Internet Dictionary Project, NDLR]. Bien que je ne sois pas multilingue, ni même bilingue moi-même, je suis conscient du fait que très peu de domaines ont une importance comparable à celle des langues et du multilinguisme. L'internet m'a permis de toucher des millions de personnes et de les aider à trouver ce qu'elles cherchaient, chose que je suis heureux de faire. (...) Dans l'ensemble, je pense que le web est important pour la sensibilisation aux langues et pour les questions culturelles. Dans quel autre endroit peut-on chercher au hasard pendant vingt minutes et trouver des informations susceptibles de vous intéresser dans trois langues différentes sinon plus? Les médias de communication rendent le monde plus petit en rapprochant les gens; je pense que le web est le premier médium - bien plus que le courrier, le télégraphe, le téléphone, la radio ou la télévision - à réellement permettre à l'usager moyen de franchir les frontières nationales et culturelles. (...) Notre monde est désormais suffisamment petit pour tenir sur un écran d'ordinateur.»

Comment Tyler voit-il l'avenir? «Je pense que l'avenir de l'internet réside dans davantage de multilinguisme, d'exploration et de compréhension multiculturelles que

nous n'en avons jamais vu. Toutefois l'internet sera seulement le médium au travers duquel l'information circule. Comme le papier qui sert de support au livre, l'internet en lui-même augmente très peu le contenu de l'information. Par contre il augmente énormément la valeur de celle-ci dans la capacité qu'il a de communiquer cette information. Dire que l'internet aiguillonne le multilinguisme est à mon sens une opinion fausse. C'est la communication qui aiguillonne le multilinguisme et l'échange multiculturel. L'internet est seulement le mode de communication le plus récent qui soit accessible aux gens plus ou moins ordinaires. L'internet a un long chemin à parcourir avant d'être omniprésent dans le monde entier, mais il est vraisemblable que lui-même ou un médium de la même lignée atteigne ce but. Les langues deviendront encore plus importantes qu'elles ne le sont quand tout le monde pourra communiquer à l'échelle de la planète (à travers le web, les discussions, les jeux, le courrier électronique, ou toute application appartenant encore au domaine de l'avenir), mais je ne sais pas si ceci mènera à un renforcement des attaches linguistiques ou à une fusion des langues jusqu'à ce qu'il n'en subsiste plus que quelques-unes ou même une seule. Une chose qui m'apparaît certaine est que l'internet sera toujours la marque de notre diversité, y compris la diversité des langues, même si cette diversité diminue. Et c'est une des choses que j'aime au sujet de l'internet, c'est un exemple à l'échelle mondiale du dicton: "Cela n'a pas vraiment disparu tant que quelqu'un s'en souvient." Et les gens se souviennent.»

Au printemps 2001, la Human-Languages Page fusionne avec le Languages Catalog (Catalogue des langues), une section de la WWW Virtual Library, pour devenir iLoveLanguages. En septembre 2003, iLoveLanguages offre 2.000 ressources linguistiques dans une centaine de langues. Quant à l'Internet Dictionary Project, faute de temps, Tyler met fin à ce projet en janvier 2007, tout en laissant les dictionnaires existants tels quels sur le web pour consultation ou téléchargement.

NetGlos

NetGlos - abrégé de The Multilingual Glossary of Internet Terminology - est lancé en 1995 à l'initiative du WorldWide Language Institute (WWLI). Il s'agit d'un projet coopératif en treize langues (allemand, anglais, chinois, croate, espagnol, français, grec, hébreu, hollandais/flamand, italien, maori, norvégien et portugais), avec la participation de nombre de traducteurs et autres professionnels des langues.

Brian King, directeur du WorldWide Language Institute, explique en septembre 1998: «Bien que l'anglais soit la langue la plus importante du web et de l'internet en général, je pense que le multilinguisme fait inévitablement partie des futures orientations du cyberespace. Voici quelques éléments qui, à mon sens, permettront que le web multilingue devienne une réalité:

1. <La popularisation des technologies de l'information>. La technologie des ordinateurs a longtemps été le seul domaine d'une élite "technicienne", à l'aise à la fois dans des langages de programmation complexes et en anglais, la langue universelle des sciences et techniques. Au départ, les ordinateurs n'ont jamais été conçus pour manier des systèmes d'écriture ne pouvant être traduits en ASCII. Il n'y avait pas de place pour autre chose que les 26 lettres de l'alphabet anglais dans un système d'encodage qui, à l'origine, ne pouvait même pas reconnaître les accents aigus et les trémas, sans parler de systèmes non alphabétiques comme le chinois. Mais la tradition

a été bouleversée, et la technologie popularisée. Des interfaces graphiques tels que Windows et Macintosh ont accéléré le processus. La stratégie de marketing de Microsoft a consisté à présenter son système d'exploitation comme facile à utiliser par le client moyen. A l'heure actuelle, cette facilité d'utilisation s'est étendue au-delà du PC vers le réseau internet, si bien que même ceux qui ne sont pas programmeurs peuvent maintenant insérer des applets Java dans leurs pages web sans comprendre une seule ligne de programmation.

2. <La compétition entre les grandes sociétés pour une part de "marché global">. L'extension de cette popularisation à l'échelon local est l'exportation des technologies de l'information dans le monde entier. La popularisation est maintenant effective à l'échelon mondial, et l'anglais n'est plus nécessairement la langue obligée de l'utilisateur. Il n'y a plus vraiment de langue indispensable, il y a les langues propres aux utilisateurs. Une chose est certaine: il n'est plus nécessaire de comprendre l'anglais pour utiliser un ordinateur, de même qu'il n'est plus nécessaire d'avoir un diplôme d'informatique. La demande des utilisateurs non anglophones - et l'effort entrepris par les sociétés de haute technologie se faisant concurrence pour obtenir les marchés mondiaux - ont fait de la localisation un secteur en expansion rapide dans le développement des logiciels et du matériel informatique. Le premier pas a été le passage de l'ASCII à l'ASCII étendu. Ceci signifie que les ordinateurs commençaient à reconnaître les accents et les symboles utilisés dans les variantes de l'alphabet anglais, symboles qui appartenaient le plus souvent aux langues européennes. Cependant une page ne pouvait être affichée que dans une seule langue à la fois.

3. <L'innovation technologique>. L'innovation la plus récente est l'Unicode. Bien qu'il soit encore en train d'évoluer et qu'il ait tout juste été incorporé dans les derniers logiciels, ce nouveau système d'encodage traduit chaque caractère en 16 octets. Alors que l'ASCII étendu à huit bits pouvait prendre en compte un maximum de 256 caractères, l'Unicode peut prendre en compte plus de 65.000 caractères uniques et il a donc la possibilité de traiter informatiquement tous les systèmes d'écriture du monde. Les instruments sont maintenant plus ou moins en place. Ils ne sont pas encore parfaits, mais on peut désormais surfer sur le web en utilisant le chinois, le japonais, le coréen, et nombre d'autres langues n'utilisant pas l'alphabet occidental. Comme l'internet s'étend à des parties du monde où l'anglais est très peu utilisé, par exemple la Chine, il est naturel que ce soit le chinois et non l'anglais qui soit utilisé. La majorité des usagers en Chine n'a pas d'autre choix que sa langue maternelle.

Une période intermédiaire précède bien sûr ce changement. Une grande partie de la terminologie technique disponible sur le web n'est pas encore traduite dans d'autres langues. Et, comme nous nous en sommes rendus compte dans NetGlos, notre glossaire multilingue de la terminologie de l'internet, la traduction de ces termes n'est pas toujours facile. Avant qu'un nouveau terme soit accepté comme le terme correct, il y a une période d'instabilité avec plusieurs candidats en compétition. Souvent un terme emprunté à l'anglais est le point de départ et, dans de nombreux cas, il est aussi le point d'arrivée. Finalement émerge un vainqueur qui est ensuite utilisé aussi bien dans les dictionnaires techniques que dans le vocabulaire quotidien de l'usager non spécialiste. La dernière version de NetGlos est la version russe, et elle devrait être disponible dans deux semaines environ [à savoir fin septembre 1998, NDLR]. Elle sera sans nul doute un excellent exemple du processus dynamique en cours pour la russification de la terminologie du web.

4. <La démocratie linguistique>. Dans un rapport de l'UNESCO du début des années 1950, l'enseignement dispensé dans sa langue maternelle était considéré comme un droit fondamental de l'enfant. La possibilité de naviguer sur l'internet dans sa langue maternelle pourrait bien être son équivalent à l'Âge de l'Information. Si l'internet doit vraiment devenir le réseau mondial qu'on nous promet, tous les usagers devraient y avoir accès sans problème de langue. Le considérer comme la chasse gardée de ceux qui, par accident historique, nécessité pratique ou privilège politique, connaissent l'anglais, est injuste à l'égard de ceux qui ne connaissent pas cette langue.

5. <Le commerce électronique>. Bien qu'un web multilingue soit souhaitable sur le plan moral et éthique, un tel idéal ne suffit pas pour en faire une réalité à vaste échelle. De même que l'utilisateur non anglophone peut maintenant avoir accès aux technologies dans sa propre langue, l'impact du commerce électronique peut constituer une force majeure qui fasse du multilinguisme la voie la plus naturelle vers le cyberespace. Les vendeurs de produits et services dans le marché virtuel mondial que devient l'internet doivent être préparés à traiter avec un monde virtuel qui soit aussi multilingue que le monde physique. S'ils veulent réussir, ils doivent s'assurer qu'ils parlent bien la langue de leurs clients!»

En ce qui concerne le WorldWide Language Institute, quelles sont les perspectives? «Comme l'existence de notre organisme est liée à l'importance attachée aux langues, je pense que son avenir sera excitant et stimulant. Mais il est impossible de pratiquer l'autosuffisance à l'égard de nos réussites et de nos réalisations. La technologie change à une allure frénétique. L'apprentissage durant toute la vie est une stratégie que nous devons tous adopter si nous voulons rester en tête et être compétitifs. C'est une tâche qui est déjà assez difficile dans un environnement anglophone. Si nous ajoutons à cela la complexité apportée par la communication dans un cyberespace multilingue et multiculturel, la tâche devient encore plus astreignante. Probablement davantage encore que par le passé, la coopération est aussi indispensable que la concurrence. Les germes d'une coopération par le biais de l'internet existent déjà. Notre projet NetGlos dépend du bon vouloir de traducteurs volontaires de nombreux pays: Canada, États-Unis, Autriche, Norvège, Belgique, Israël, Portugal, Russie, Grèce, Brésil, Nouvelle-Zélande, etc. À mon avis, les centaines de visiteurs qui consultent quotidiennement les pages de NetGlos constituent un excellent témoignage du succès de ce type de relations de travail. Les relations de coopération s'accroîtront encore à l'avenir, mais pas nécessairement sur la base du volontariat.»

Logos

Fondé en 1979 à Modène (Italie) par Rodrigo Vergara, Logos est une société de traduction offrant des services dans 35 langues en 1997, avec 300 traducteurs travaillant sur place et un réseau mondial de 2.500 traducteurs travaillant en free-lance. La moyenne de production est de 200 textes par jour.

Fin 1997, Logos décide de mettre tous ses outils professionnels en accès libre sur le web. Le Logos Dictionary est un dictionnaire multilingue de 7,5 millions d'entrées. La Wordtheque est une base de données multilingue de 328 millions de mots, constituée à partir de milliers de traductions, notamment des romans et des documents techniques. La recherche dans la Wordtheque est possible par langue, par mot, par auteur ou par

titre. Linguistic Resources offre un point d'accès unique à 553 glossaires. L'Universal Conjugator propose des tableaux de conjugaison dans 17 langues.

Annie Kahn, journaliste au quotidien *Le Monde*, publie le 7 décembre 1997 un article *Les mots pour le dire*, dans lequel Robert Vergara relate: «Nous voulions que nos traducteurs aient tous accès aux mêmes outils de traduction. Nous les avons donc mis à leur disposition sur internet, et tant qu'à faire nous avons ouvert le site au public. Cela nous a rendus très populaires, nous a fait beaucoup de publicité. L'opération a drainé vers nous de nombreux clients, mais aussi nous a permis d'étoffer notre réseau de traducteurs grâce aux contacts établis à la suite de cette initiative.»

Annie Kahn explique dans le même article: «Le site de Logos est beaucoup plus qu'un dictionnaire ou qu'un répertoire de liens vers d'autres dictionnaires en ligne. L'un des piliers du système est un logiciel de recherche documentaire fonctionnant sur un corpus de textes littéraires disponibles gratuitement sur internet. Lorsque l'on recherche la définition ou la traduction d'un mot, "didactique" par exemple, on trouve non seulement le résultat recherché, mais aussi une phrase d'une œuvre littéraire utilisant ce mot (en l'occurrence, un essai de Voltaire). Un simple clic permet d'accéder au texte intégral de l'œuvre ou de commander le livre grâce à un partenariat avec Amazon.com, le libraire en ligne bien connu. Il en est de même avec les traductions étrangères. Si aucun texte utilisant ce mot n'a été trouvé, le système fonctionne alors comme un moteur de recherche et renvoie aux sites web concernant ce mot. Pour certains termes, il est proposé d'en entendre la prononciation. Si une traduction manque, le système fait un appel au peuple. A chacun d'enrichir la base, les traducteurs de l'entreprise valident ensuite les traductions proposées.»

En 2007, la Wordtheque, devenue la Logos Library, comprend 710 millions de termes. Conjugation of Verbs, devenu l'Universal Conjugator, propose des tableaux de conjugaison dans 36 langues. Et Linguistic Resources offre un point d'accès unique à 1.215 glossaires.

Dictionnaires de langues

Dictionnaires imprimés en ligne
Répertoires de dictionnaires
yourDictionary.com
Le Grand dictionnaire terminologique
Eurodicautom et IATE
WordReference.com

Dictionnaires imprimés en ligne

Le premier dictionnaire de langue française en accès libre est le *Dictionnaire universel francophone* en ligne, qui répertorie 45.000 mots et 116.000 définitions tout en présentant «sur un pied d'égalité, le français dit "standard" et les mots et expressions en français tel qu'on le parle sur les cinq continents».

Issu de la collaboration entre Hachette et l'AUPELF-UREF (devenu depuis l'AUF: Agence universitaire de la Francophonie), il correspond à la partie «noms communs» du dictionnaire imprimé disponible chez Hachette. L'équivalent pour la langue anglaise est le site Merriam-Webster OnLine, qui donne librement accès au Collegiate Dictionary et au Collegiate Thesaurus.

En mars 2000, les 20 volumes de l'Oxford English Dictionary (OED) sont mis en ligne par l'Oxford University Press (OUP). La consultation du site est payante. Le dictionnaire bénéficie d'une mise à jour trimestrielle d'environ 1.000 entrées nouvelles ou révisées.

Deux ans après cette première expérience, en mars 2002, l'Oxford University Press met en ligne l'Oxford Reference Online (ORO), une vaste encyclopédie conçue directement pour le web et consultable elle aussi sur abonnement payant. Avec 60.000 pages et un million d'entrées, elle représente l'équivalent d'une centaine d'ouvrages de référence.

Répertoires de dictionnaires

Le répertoire Dictionnaires électroniques est un excellent répertoire établi par la section française des Services linguistiques centraux (SLC-f) de l'Administration fédérale suisse. Cette liste très complète de dictionnaires monolingues (allemand, anglais, espagnol, français, italien), bilingues et multilingues est complétée par des répertoires d'abréviations et acronymes et des répertoires géographiques, essentiellement des atlas.

Marcel Grangier, responsable de la section française des Services linguistiques centraux, écrit en janvier 1999: «Travailler sans internet est devenu tout simplement impossible: au-delà de tous les outils et commodités utilisés (messagerie électronique, consultation de la presse électronique, activités de services au profit de la profession des traducteurs), internet reste pour nous une source indispensable et inépuisable d'informations dans ce que j'appellerais le "secteur non structuré" de la toile. Pour illustrer le propos, lorsqu'aucun site comportant de l'information organisée ne fournit de réponse à un problème de traduction, les moteurs de recherche permettent dans la plus grande partie des cas de retrouver le chaînon manquant quelque part sur le réseau.»

Comment voit-il l'avenir? «La multiplication des langues présentes sur internet est inévitable, et ne peut que bénéficier aux échanges multiculturels. Pour que ces échanges prennent place dans un environnement optimal, il convient encore de développer les outils qui amélioreront la compatibilité. La gestion complète des diacritiques ne constitue qu'un exemple de ce qui peut encore être entrepris.»

Quelques années après, le répertoire Dictionnaires électroniques rejoint le site de la Conférence des Services de traduction des États européens (CST).

yourDictionary.com

Robert Beard, professeur de langues à la Bucknell University (États-Unis), crée d'abord en 1995 A Web of Online Dictionaries (Un web de dictionnaires en ligne), qui est un répertoire de dictionnaires en ligne (avec 800 liens en automne 1998) dans de nombreuses langues, auquel s'ajoutent d'autres sections: dictionnaires multilingues, dictionnaires anglophones spécialisés, thésauri et vocabulaires, grammaires en ligne, et enfin outils linguistiques pour non spécialistes.

Robert Beard écrit en septembre 1998: «On a d'abord craint que le web représente un danger pour le multilinguisme, étant donné que l'HTML et d'autres langages de programmation sont basés sur l'anglais et qu'on trouve tout simplement plus de sites web en anglais que dans toute autre langue. Cependant, les sites web que je gère montrent que le multilinguisme est très présent et que le web peut en fait permettre de préserver des langues menacées de disparition. Je propose maintenant des liens vers des dictionnaires dans 150 langues et des grammaires dans 65 langues. De plus, ceux qui développent les logiciels de navigation manifestent une attention nouvelle pour la diversité des langues dans le monde, ce qui favorisera la présence d'un nombre encore plus grand de sites web dans différentes langues.»

Robert Beard co-fonde ensuite le portail yourDictionary.com, qui intègre son site précédent, avec mise en ligne du portail en février 2000.

Il écrit en janvier 2000: «Nos nouvelles idées sont nombreuses. Nous projetons de travailler avec l'Endangered Language Fund aux États-Unis et en Grande-Bretagne pour rassembler des fonds pour cette fondation et nous publierons les résultats sur notre site. Nous aurons des groupes de discussion et des bulletins d'information sur les langues. Il y aura des jeux de langue destinés à se distraire et à apprendre les bases de la linguistique. La page Linguistic Fun [qui propose des éléments de linguistique pour les non initiés, NDLR] deviendra un journal en ligne avec des extraits courts, intéressants et même amusants dans différentes langues, choisis par des experts du monde entier. (...) Si l'anglais domine encore le web, on voit s'accentuer le développement de sites monolingues et non anglophones du fait des solutions variées apportées aux problèmes de caractères.»

En septembre 2003, yourDictionary.com, devenu un portail de référence, répertorie plus de 1.800 dictionnaires dans 250 langues, ainsi que de nombreux outils linguistiques: vocabulaires, grammaires, glossaires, méthodes de langues, etc. En avril 2007, le répertoire comprend 2.500 dictionnaires et grammaires dans 300 langues.

Soucieux de servir toutes les langues sans exception, le portail propose comme prévu

l'Endangered Language Repository, une section spécifique consacrée aux langues menacées. «Les langues menacées sont essentiellement des langues non écrites, écrit Robert Beard en janvier 2000. Un tiers seulement des quelques 6.000 langues existant dans le monde sont à la fois écrites et parlées. Je ne pense pourtant pas que le web va contribuer à la perte de l'identité des langues et j'ai même le sentiment que, à long terme, il va renforcer cette identité. Par exemple, de plus en plus d'Indiens d'Amérique contactent des linguistes pour leur demander d'écrire la grammaire de leur langue et de les aider à élaborer des dictionnaires. Pour eux, le web est un instrument à la fois accessible et très précieux d'expression culturelle.»

Le Grand Dictionnaire terminologique

Le Grand dictionnaire terminologique (GDT) est une initiative majeure de l'Office québécois de la langue française (OQLF). C'est en effet la première fois qu'un organisme propose une base terminologique aussi importante en accès libre sur le web, en septembre 2000. Le GDT est précédé par Le Signet, une base terminologique pour les technologies de l'information, dont les 10.000 fiches bilingues français-anglais sont ensuite intégrées au GDT.

Le GDT est un dictionnaire bilingue français-anglais de 3 millions de termes appartenant au vocabulaire industriel, scientifique et commercial. Sa mise en ligne est le résultat d'un partenariat entre l'OQLF, auteur du dictionnaire, et Semantix, société spécialisée dans les solutions logicielles linguistiques. Événement célébré par de très nombreux linguistes, cette mise en ligne est un succès. Dès le premier mois, le GDT est consulté par 1,3 million de personnes, avec des pointes de 60.000 requêtes quotidiennes. La gestion de la base est ensuite assurée par Convera Canada. En février 2003, les requêtes sont au nombre de 3,5 millions par mois. Une nouvelle version du GDT est mise en ligne en mars 2003. Sa gestion est désormais assurée par l'OQLF lui-même, et non plus par une société prestataire.

Eurodicautom et IATE

Géré par le service de traduction de la Commission européenne, Eurodicautom est une base terminologique multilingue de termes économiques, scientifiques, techniques et juridiques permettant de combiner entre elles les onze langues officielles de l'Union européenne (allemand, anglais, danois, espagnol, finnois, français, grec, hollandais, italien, portugais, suédois), ainsi que le latin, avec une moyenne de 120.000 consultations par jour en 2003.

Fin 2003, Eurodicautom annonce son intégration dans une base terminologique plus vaste regroupant les bases terminologiques de plusieurs institutions de l'Union européenne, notamment EUTERPE, la base du Parlement européen, et TIS, la base du Conseil de l'Union européenne. Cette nouvelle base traiterait non plus douze langues mais une vingtaine, du fait de l'élargissement de l'Union européenne prévu l'année suivante pour intégrer plusieurs pays d'Europe de l'Est.

Un projet de base terminologique commune est évoqué dès 1999 afin de renforcer la coopération inter-institutionnelle. Les partenaires de ce projet sont le Parlement européen, le Conseil de l'Union européenne, la Commission européenne, la Cour de justice, la Cour des comptes européenne, le Comité économique et social européen, le Comité des régions, la Banque européenne d'investissement, la Banque centrale

européenne et enfin le Centre de traduction des organes de l'Union européenne.

La nouvelle base terminologique voit le jour au printemps 2004 sous le nom de IATE (Inter-Active Terminology for Europe), d'abord pour un usage interne dans les institutions de l'Union européenne avant de migrer sur le web en juin 2007 en tant que service public, avec 1,4 million d'entrées dans les 23 langues officielles de l'Union européenne, plus le latin. L'Union européenne est en effet passée de 15 à 25 pays membres en mai 2004, pour atteindre 27 pays membres en janvier 2007, d'où la nécessité de 23 langues officielles au lieu des 11 langues officielles présentes dans Eurodicautom.

Le site web de IATE est administré par le Centre de traduction des organes de l'Union européenne à Luxembourg (capitale du pays du même nom), pour le compte des partenaires du projet. Comme expliqué dans la brochure de IATE, elle-même disponible en 23 langues, «les termes sont introduits dans la base de données par les terminologues et les traducteurs de l'Union européenne sur la base des informations fournies par les traducteurs, les administrateurs, les juristes-linguistes, les experts et d'autres sources fiables.» En 2009, IATE comprend 8,4 millions de termes, dont 540.000 abréviations et 130.000 expressions.

WordReference.com

Le site WordReference.com est lancé en 1999 par Michael Kellogg pour offrir des dictionnaires bilingues gratuits en ligne. En mars 2010, Michael relate sur son site: «L'internet a été un incroyable outil ces dernières années pour rassembler des gens du monde entier. L'un des principaux obstacles à cela reste bien entendu la langue. Le contenu de l'internet est pour une grande part en anglais et de très nombreux usagers lisent ces pages alors que l'anglais est leur deuxième langue et non leur langue maternelle. De par mes propres expériences avec la langue espagnole, je sais que de nombreux lecteurs comprennent une grande partie de ce qu'ils lisent, mais pas la totalité.

J'ai débuté ce site en 1999 pour procurer des dictionnaires bilingues gratuits en ligne et d'autres outils pour tous sur l'internet. Depuis, le site s'est progressivement développé pour devenir l'un des sites de dictionnaires en ligne les plus utilisés, et le principal dictionnaire en ligne pour les paires de langues anglais-espagnol, anglais-français, anglais-italien, espagnol-français et espagnol-portugais. Ce site est toujours classé sans interruption parmi les 500 sites les plus visités du web. Aujourd'hui, je suis heureux de continuer à améliorer ces dictionnaires, les autres outils linguistiques du site et les forums de langues. J'ai vraiment plaisir à créer de nouvelles fonctionnalités pour rendre ce site de plus en plus utile.»

Les dictionnaires les plus populaires sont le dictionnaire espagnol (espagnol-anglais et anglais-espagnol), le dictionnaire français et le dictionnaire italien. On trouve aussi un dictionnaire allemand, un dictionnaire russe et un dictionnaire monolingue anglais. Des tableaux de conjugaison sont disponibles pour l'espagnol, le français et l'italien.

Pour l'anglais, on trouve également des dictionnaires de l'anglais vers les langues suivantes: arabe, chinois, coréen, grec, japonais, polonais, portugais, roumain, tchèque et turc, et vice versa.

Pour l'espagnol, en plus des deux dictionnaires d'Espasa Calpe et d'Oxford complétés par le supplément propre à WordReference.com, on peut consulter un dictionnaire monolingue espagnol, un dictionnaire espagnol de synonymes, un dictionnaire espagnol-français et un dictionnaire espagnol-portugais.

Pour le français et l'italien, outre les dictionnaires d'Oxford, WordReference.com propose deux dictionnaires qui lui sont propres, à savoir un dictionnaire français-anglais de 250.000 termes et un dictionnaire italien-anglais de 200.000 termes.

WordReference.com offre également des forums linguistiques très actifs et de qualité. Si les usagers ont une question sur un problème linguistique donné, ils peuvent faire une recherche dans les centaines de milliers de questions précédentes, avant de poser leur propre question dans l'un des forums si nécessaire, pour être aidés par des gens des quatre coins du monde.

WordReference Mini est une version miniature du site qui permet son intégration dans d'autres sites, par exemple des sites d'apprentissage de langues.

Une version pour appareil mobile est également disponible pour plusieurs dictionnaires: anglais-espagnol, espagnol-anglais, anglais-français, français-anglais, anglais-italien, italien-anglais, avec d'autres paires de langues à venir.

L'apprentissage des langues

Quelques expériences
Des outils pour les enseignants
La LINGUIST List

Quelques expériences

Maria Victoria Marinetti, de nationalité mexicaine, est titulaire d'un doctorat en ingénierie. Depuis son installation en France, elle est professeur d'espagnol dans plusieurs entreprises du bassin annécien, en Haute-Savoie, et également traductrice.

Elle raconte en août 1999: «J'ai accès à un nombre important d'informations au niveau mondial, ce qui est très intéressant pour moi. J'ai également la possibilité de transmettre ou de recevoir des fichiers, dans un va-et-vient d'information constant. L'internet me permet de recevoir ou d'envoyer des traductions générales ou techniques du français vers l'espagnol et vice versa, ainsi que des textes espagnols corrigés. Dans le domaine technique ou chimique, je propose une aide technique, ainsi que des informations sur l'exportation d'équipes de haute technologie vers le Mexique ou d'autres pays d'Amérique latine.»

Elle ajoute en août 2001: «Depuis notre premier entretien, j'utilise beaucoup l'internet pour des échanges avec ma famille au Mexique et avec mes amis un peu partout dans le monde. C'est un outil de communication rapide, agréable et fantastique pour moi. Par contre, pour l'utilisation d'internet comme outil de télétravail, très peu d'entreprises ont le matériel et l'expérience nécessaires pour échanger des données dans le travail quotidien, notamment par la voix et l'image (par exemple pour la formation ou les conférences par l'internet). Pour ma part, je rencontre ce problème car je souhaite proposer une téléformation en langue espagnole, en utilisant la voix et l'image. Mais mes entreprises clientes ne sont pas habituées à utiliser ces moyens de communication malgré leur caractère pratique (pas de déplacements à faire) et malgré la fiabilité accrue de ces nouveaux moyens de communication par l'internet. En conclusion, les sociétés de conseil informatique ont encore beaucoup à faire pour familiariser les entreprises à l'utilisation des nouvelles technologies liées aux transferts de données par l'internet.»

Robert Beard, professeur de langues et créateur du portail yourDictionary.com, écrit en septembre 1998: «En tant que professeur de langues, je pense que le web présente une pléthore de nouvelles ressources disponibles dans la langue étudiée, de nouveaux instruments d'apprentissage (exercices interactifs Java et Shockwave) et de test, qui sont à la disposition des étudiants quand ceux-ci en ont le temps ou l'envie, 24 heures par jour et 7 jours par semaine. Aussi bien pour mes collègues que pour moi, et bien sûr pour notre établissement, l'internet nous permet aussi de publier pratiquement sans limitation.»

Comment voit-il l'avenir? «L'internet nous offrira tout le matériel pédagogique dont nous pouvons rêver, y compris des notes de lecture, exercices, tests, évaluations et exercices interactifs plus efficaces que par le passé, parce que reposant davantage sur la notion de communication.»

Une autre expérience est celle de Russon Wooldridge, professeur au département des études françaises de l'Université de Toronto (Canada), qui relate en février 2001: «Tout mon enseignement exploite au maximum les ressources d'internet (le web et le courriel): les deux lieux communs d'un cours sont la salle de classe et le site du cours, sur lequel je mets tous les matériaux des cours. Je mets toutes les données de mes recherches des vingt dernières années sur le web (réédition de livres, articles, textes intégraux de dictionnaires anciens en bases de données interactives, de traités du 16e siècle, etc.). Je publie des actes de colloques, j'édite un journal, je collabore avec des collègues français, mettant en ligne à Toronto ce qu'ils ne peuvent pas publier en ligne chez eux. En mai 2000 j'ai organisé à Toronto un colloque international sur "Les études françaises valorisées par les nouvelles technologies". (...) Je me rends compte que sans internet mes activités seraient bien moindres, ou du moins très différentes de ce qu'elles sont actuellement. Donc je ne vois pas l'avenir sans.»

Des outils pour les enseignants

Depuis ses débuts en 1989, le Computer in Teaching Initiative (CTI) Centre for Modern Languages est un centre inclus dans l'Institut des langues de l'Université d'Hull (Royaume-Uni) et vise à promouvoir l'utilisation des ordinateurs dans l'apprentissage et l'enseignement des langues. Connu sous le nom de CTI Centre, il procure des informations sur la manière dont l'apprentissage des langues assisté par ordinateur peut être effectivement intégré à des cours existants, et il offre un soutien aux professeurs qui utilisent - ou souhaitent utiliser - l'informatique dans l'enseignement qu'ils dispensent.

June Thompson, responsable du CTI Centre, écrit en décembre 1998: «Avec l'internet, on a la possibilité de favoriser l'utilisation des langues étrangères, et notre organisation ne soutient absolument pas la suprématie de l'anglais en tant que langue de l'internet. L'utilisation de l'internet a apporté une nouvelle dimension à notre tâche qui consiste à soutenir les professeurs de langue dans l'utilisation de la technologie correspondante. Je pense que, dans un avenir proche, l'utilisation de supports linguistiques sur l'internet va continuer à se développer en même temps que d'autres activités liées aux technologies, par exemple l'utilisation de CD-ROM - certains établissements n'ont pas suffisamment de matériel informatique en réseau. A l'avenir, il me semble que l'utilisation de l'internet jouera un rôle plus grand, mais seulement si ces activités sont à caractère pédagogique. Notre organisme travaille étroitement avec le WELL, qui se consacre à ces problèmes.»

Le WELL (Web Enhanced Language Learning) est un projet britannique mené à bien entre 1997 et 2000 pour donner accès à des ressources en ligne de qualité dans douze langues différentes. Sélectionnées et décrites par des experts, ces ressources sont complétées par des informations et des exemples sur la manière de les utiliser pour l'enseignement ou l'apprentissage d'une langue.

Ce projet est l'œuvre de l'association EUROCALL (European Association for Computer-Assisted Language Learning), qui regroupe des professionnels de l'enseignement des langues exerçant en Europe et dans le monde entier. Ses objectifs sont de favoriser l'utilisation des langues étrangères en Europe, encourager une vision européenne de l'utilisation des technologies pour l'apprentissage des langues, et enfin promouvoir la création et la diffusion d'un matériel de qualité.

Un autre projet d'EUROCALL est CAPITAL (Computer-Assisted Pronunciation Investigation Teaching and Learning), qui regroupe des chercheurs et praticiens souhaitant utiliser l'informatique dans ce domaine.

La LINGUIST List

Gérée par l'Eastern Michigan University et la Wayne State University, deux universités des États-Unis, la LINGUIST List est une liste de diffusion à destination des linguistes. Les messages reçus sont classés dans diverses rubriques: profession (conférences, associations linguistiques, programmes, etc.), recherche et soutien à la recherche (articles, résumés de mémoires, projets, bibliographies, dossiers, textes), publications, pédagogie, ressources linguistiques (langues, familles linguistiques, dictionnaires, informations régionales) et soutien informatique (polices de caractères et logiciels). La LINGUIST List propose aussi un centre de documentation virtuel, dénommé Virtual Library.

Helen Dry, modératrice de la LINGUIST List, explique en août 1998: «La LINGUIST List, que je modère, a pour politique d'accepter les informations dans toutes les langues, puisque c'est une liste pour linguistes. Nous ne souhaitons cependant pas que le message soit publié dans plusieurs langues, tout simplement à cause de la charge de travail que cela représenterait pour notre personnel de rédaction (nous ne sommes pas une liste fourre-tout, mais une liste modérée: avant d'être publié, chaque message est classé par nos étudiants-rédacteurs dans une section comprenant des messages du même type). Notre expérience nous montre que pratiquement tout le monde choisit de publier en anglais. Mais nous relions ces informations à un système de traduction qui présente nos pages dans cinq langues différentes. Ainsi un abonné ne lit LINGUIST en anglais que s'il le souhaite. Nous essayons aussi d'avoir au moins un étudiant-éditeur qui soit réellement multilingue, afin que les lecteurs puissent correspondre avec nous dans d'autres langues que l'anglais.»

Les langues minoritaires

La liste de Caoimhín
Le site Windows on Haiti

La liste de Caoimhín

Contrairement aux clichés véhiculés dans les médias, l’internet ne favorise pas forcément l'hégémonie de l'anglais et n'entraîne pas la disparition des langues minoritaires. L'internet peut au contraire contribuer à protéger ces langues, s'il existe une volonté politique et culturelle dans ce sens. En témoigne l'expérience de Caoimhín.

Caoimhín Ó Donnaíle est professeur d’informatique à l’Institut Sabhal Mór Ostaig, situé sur l’île de Skye, en Écosse. Il dispense ses cours en gaélique écossais. Il est également le webmestre du site de l’institut, qui est bilingue anglais-gaélique et qui se trouve être la principale source d’information mondiale sur le gaélique écossais. Sur ce site, il tient à jour la page European Minority Languages, une liste elle aussi bilingue anglais-gaélique, avec classement par ordre alphabétique de langues et par famille linguistique.

Interviewé en août 1998, Caoimhín raconte: «L'internet a contribué et contribuera au développement fulgurant de l'anglais comme langue mondiale. L'internet peut aussi grandement aider les langues minoritaires. Ceci ne se fera pas tout seul, mais seulement si les gens choisissent de défendre une langue. Le web est très utile pour dispenser des cours de langues, et la demande est grande.»

Près de trois ans plus tard, en mai 2001, il ajoute: «Nos étudiants utilisent un correcteur d’orthographe en gaélique et une base terminologique en ligne en gaélique. (...) Il est maintenant possible d’écouter la radio en gaélique (écossais et irlandais) en continu sur l’internet partout dans le monde. Une réalisation particulièrement importante a été la traduction en gaélique du navigateur Opera. C’est la première fois qu’un logiciel de cette taille est disponible en gaélique.»

La langue gaélique est promue par toute une communauté linguistique. Mais qu'en est-il des langues menacées? «En ce qui concerne l’avenir des langues menacées, l’internet accélère les choses dans les deux sens. Si les gens ne se soucient pas de préserver les langues, l’internet et la mondialisation qui l’accompagne accéléreront considérablement la disparition de ces langues. Si les gens se soucient vraiment de les préserver, l’internet constituera une aide irremplaçable.»

Le site Windows on Haiti

Guy Antoine, créateur de Windows on Haiti, site de référence sur la langue haïtienne, relate en novembre 1999: «J’ai fait de la promotion du kreyòl (créole haïtien) une cause personnelle, puisque cette langue est le principal lien unissant tous les Haïtiens, malgré l’attitude dédaigneuse d’une petite élite haïtienne - à l’influence disproportionnée - vis-à-vis de l’adoption de normes pour l’écriture du kreyòl et le soutien de la publication de livres et d’informations officielles dans cette langue. A titre d’exemple, il y avait récemment dans la capitale d’Haïti un Salon du livre de deux semaines, à qui on avait donné le nom de "Livres en folie". Sur les 500 ouvrages d’auteurs haïtiens présentés lors du salon, il y en avait une vingtaine en kreyòl, ceci dans le cadre de la campagne

insistante que mène la France pour célébrer la Francophonie dans ses anciennes colonies. A Haïti cela se passe relativement bien, mais au détriment direct de la Créolophonie.

En réponse à l'attitude de cette minorité haïtienne, j'ai créé sur mon site Windows on Haiti deux forums de discussion exclusivement en kreyòl. Le premier forum regroupe des discussions générales sur toutes sortes de sujets, mais en fait ces discussions concernent principalement les problèmes socio-politiques qui agitent Haïti. Le deuxième forum est uniquement réservé aux débats sur les normes d'écriture du kreyòl. Ces débats sont assez animés, et un certain nombre d'experts linguistiques y participent. Le caractère exceptionnel de ces forums est qu'ils ne sont pas académiques. Je n'ai trouvé nulle part ailleurs sur l'internet un échange aussi spontané et aussi libre entre des experts et le grand public pour débattre dans une langue donnée des attributs et des normes de la même langue.»

En juin 2001, Guy Antoine rejoint l'équipe dirigeante de Mason Integrated Technologies, une société dont l'objectif est de créer des outils permettant l'accessibilité des documents publiés dans des langues dites minoritaires. «Étant donné l'expérience de l'équipe en la matière, nous travaillons d'abord sur le créole haïtien (kreyòl), qui est la seule langue nationale d'Haïti, et l'une des deux langues officielles (l'autre étant le français). Cette langue ne peut guère être considérée comme une langue minoritaire dans les Caraïbes puisqu'elle est parlée par huit à dix millions de personnes.»

Des encyclopédies multilingues

Des projets précurseurs
Wikipédia

Des projets précurseurs

Robert Beard, professeur de langues et co-fondateur du portail yourDictionary.com, écrit en septembre 1998: «Le web sera une encyclopédie du monde faite par le monde pour le monde. Il n'y aura plus d'informations ni de connaissances utiles qui ne soient pas disponibles, si bien que l'obstacle principal à la compréhension internationale et interpersonnelle et au développement personnel et institutionnel sera levé. Il faudrait une imagination plus débordante que la mienne pour prédire l'effet de ce développement sur l'humanité.»

Les premières grandes encyclopédies en ligne apparaissent en décembre 1999 avec WebEncyclo et l'Encyclopaedia Universalis en langue française et Britannica.com en langue anglaise.

WebEncyclo, publié par les éditions Atlas, est la première grande encyclopédie francophone en accès libre. La recherche est possible par mots-clés, thèmes, médias (à savoir les cartes, liens internet, photos ou illustrations) et idées. Un appel à contribution incite les spécialistes d'un sujet donné à envoyer des articles, qui sont regroupés dans la section WebEncyclo contributif. Après avoir été libre, l'accès est ensuite soumis à une inscription préalable gratuite.

La version web de l'Encyclopaedia Universalis est mise en ligne à la même date, soit un ensemble de 28.000 articles signés par 4.000 auteurs. Si la consultation est payante sur la base d'un abonnement annuel, de nombreux articles sont en accès libre.

Le site Britannica.com est la première grande encyclopédie anglophone en accès libre. Le site web propose l'équivalent numérique des 32 volumes de la 15e édition de l'Encyclopaedia Britannica, parallèlement à la version imprimée et à la version CD-ROM, toutes deux payantes. Le site offre aussi une sélection d'articles issus de 70 magazines, un guide des meilleurs sites, un choix de livres, etc., le tout étant accessible à partir d'un moteur de recherche unique. En septembre 2000, le site fait partie des cent sites les plus visités au monde. En juillet 2001, la consultation devient payante sur la base d'un abonnement annuel ou mensuel. Fin 2008, Britannica.com annonce l'ouverture prochaine de son site à des contributeurs extérieurs, avec inscription obligatoire pour écrire et modifier des articles.

Wikipédia

Issu du terme hawaïen «wiki» (qui signifie: vite, rapide), un wiki est un site web permettant à plusieurs utilisateurs de collaborer en ligne sur un même projet. A tout moment, ces utilisateurs peuvent contribuer à la rédaction du contenu, modifier ce contenu et l'enrichir en permanence. Le wiki est utilisé par exemple pour créer et gérer des dictionnaires, des encyclopédies ou encore des sites d'information sur un sujet donné. Le programme présent derrière l'interface d'un wiki est plus ou moins élaboré. Un programme simple gère du texte et des hyperliens. Un programme élaboré permet

d'inclure des images, des graphiques, des tableaux, etc. L'encyclopédie wiki la plus connue est Wikipédia.

Fondée en janvier 2001 à l'initiative de Jimmy Wales et de Larry Sanger (Larry quitte ensuite l'équipe), Wikipédia est une encyclopédie gratuite écrite collectivement et dont le contenu est librement réutilisable. Elle est immédiatement très populaire. Sans publicité et financée par des dons, cette encyclopédie coopérative est rédigée par des milliers de volontaires - appelés Wikipédiens, et qui s'inscrivent sous un pseudonyme - avec possibilité de corriger et compléter les articles, aussi bien les leurs que ceux d'autres contributeurs. Les articles restent la propriété de leurs auteurs, et leur libre utilisation est régie par la licence GFDL (GNU free documentation license).

En décembre 2004, Wikipédia compte 1,3 million d'articles rédigés dans une centaine de langues par 13.000 contributeurs. En décembre 2006, Wikipédia compte 6 millions d'articles dans 250 langues, et devient l'un de dix sites les plus visités du web. En mai 2007, Wikipédia compte 7 millions d'articles dans 192 langues, dont 1,8 million en anglais, 589.000 en allemand, 500.000 en français, 260.000 en portugais et 236.000 en espagnol. En 2009, Wikipédia fait partie des cinq sites les plus visités du web.

Fondée en juin 2003, la Wikimedia Foundation gère non seulement Wikipédia mais aussi Wiktionary, un dictionnaire et thésaurus multilingue lancé en décembre 2002, puis Wikibooks (livres et manuels en cours de rédaction) lancé en juin 2003, auxquels s'ajoutent ensuite Wikiquote (répertoire de citations), Wikisource (textes appartenant au domaine public), Wikimedia Commons (sources multimédia), Wikispecies (répertoire d'espèces animales et végétales), Wikinews (site d'actualités) et enfin Wikiversity (matériel d'enseignement), lancé en août 2006.

Localisation et internationalisation

Peter Raggett, sous-directeur (puis directeur) du Centre de documentation et d'information (CDI) de l'OCDE (Organisation de coopération et de développement économiques), écrit en août 1999: «Je pense qu'il appartient aux organisations et sociétés européennes d'offrir des sites web si possible en trois ou quatre langues. À l'heure de la mondialisation et du commerce électronique, les sociétés ont un marché potentiel sur plusieurs pays à la fois. Permettre aux usagers francophones, germanophones ou nippons de consulter un site web aussi facilement que les usagers anglophones donnera une plus grande compétitivité à une firme donnée.»

«Vers la communication sur internet dans toutes les langues...», tel est le sous-titre de la page d'accueil de Babel, un projet conjoint d'Alis Technologies et de l'Internet Society lancé en 1997 dans l'optique d'une internationalisation de l'internet. À cette date, le site multilingue de Babel (en anglais, allemand, espagnol, français, italien, portugais et suédois) propose deux grands secteurs: (a) un secteur langues, avec trois sections: langues du monde, glossaire typographique et linguistique, Francophonie; (b) un secteur internet et multilinguisme, avec deux sections: développer votre site web multilingue, et codage des écritures du monde. Babel propose aussi la page *Palmarès des langues de la toile*, qui est la première à donner la répartition réelle des langues sur le réseau.

Bill Dunlap est le fondateur de Euro-Marketing Associates, une société de conseil en marketing qu'il lance en 1985 à Paris et San Francisco. En 1995, il restructure cette société en service de conseil en ligne dénommé Global Reach, le but étant de promouvoir les sites web des entreprises dans d'autres pays, afin d'attirer plus de visiteurs, et donc d'augmenter les ventes. Cette méthode comprend la traduction d'un site web dans plusieurs langues, la promotion active du site, et enfin l'accroissement de la fréquentation locale au moyen de bandeaux publicitaires ciblés.

Bill Dunlap explique en décembre 1998: «Il y a très peu de gens aux États-Unis qui sont intéressés de communiquer dans plusieurs langues. Pour la plupart, ils pensent encore que le monde entier parle anglais. Par contre, en Europe, les pays sont petits, si bien que, depuis des siècles, une perspective internationale est nécessaire. Depuis 1981, début de mon activité professionnelle, j'ai été impliqué dans la venue de sociétés américaines en Europe. Ceci est pour beaucoup un problème de langue, puisque leurs informations commerciales doivent être disponibles dans les langues européennes pour être prises en compte en Europe. Comme le web est devenu populaire en 1995, j'ai donné à ces activités une dimension "en ligne", et j'en suis venu à promouvoir le cybercommerce européen auprès de mes compatriotes américains. (...)

Promouvoir un site est aussi important que le créer, sinon plus. On doit être préparé à utiliser au moins autant de temps et d'argent à promouvoir son site qu'on en a passé à l'origine à le créer. Le programme Global Reach permet de promouvoir un site dans des pays non anglophones, afin d'atteindre une clientèle plus large... et davantage de ventes. Une société a de nombreuses bonnes raisons de considérer sérieusement le marché international. Global Reach est pour elle le moyen d'étendre son site web à de nombreux pays, de le présenter à des visiteurs en ligne dans leur propre langue, et de

pénétrer le réseau de commerce en ligne présent dans ces pays.»

Bill Dunlap ajoute en juillet 1999: «Une fois que la page d'accueil d'un site est disponible en plusieurs langues, l'étape suivante est le développement du contenu dans chaque langue. Un webmestre notera quelles langues attirent plus de visiteurs (et donc plus de ventes) que d'autres. Ce seront donc dans ces langues que débutera une campagne de promotion multilingue sur le web. Parallèlement, il est toujours bon de continuer à augmenter le nombre de langues dans lesquelles un site web est disponible. Au début, seule la page d'accueil traduite en plusieurs langues suffit, mais ensuite il est souhaitable de développer un véritable secteur pour chaque langue.»

Le World Wide Web Consortium (W3C) est un consortium industriel international fondé en 1994 pour développer les protocoles communs du web. Le site du W3C propose notamment une section Internationalization/Localization, qui donne une définition des protocoles utilisés: HTML (hypertext markup language), jeux (de base) de caractères, nouveaux attributs, HTTP (hypertext transfer protocol), négociation de la langue, URL (uniform resource locator) et autres identificateurs incluant des caractères non ASCII (American standard code for information interchange). Le site propose aussi des conseils pour créer un site multilingue.

La traduction assistée par ordinateur

L'internet étant une source d'information à vocation mondiale, il semble indispensable d'augmenter fortement les activités de traduction. Auteur des *Chroniques de Cybérie*, une chronique hebdomadaire en ligne des actualités du réseau, Jean-Pierre Cloutier déplore en août 1999 «qu'il se fasse très peu de traductions des textes et essais importants qui sont publiés sur le web, tant de l'anglais vers d'autres langues que l'inverse. (...) La nouveauté d'internet dans les régions où il se déploie présentement y suscite des réflexions qu'il nous serait utile de lire. À quand la traduction des penseurs hispanophones et autres de la communication?»

Créé à Amsterdam (Pays-Bas) par la firme Vorontsoff, Wesseling & Partners, Aquarius est le premier répertoire non commercial de traducteurs et interprètes. En novembre 1998, il comprend 6.100 traducteurs, 800 sociétés de traduction, 91 domaines d'expertise et 369 combinaisons de langues. Le site permet de localiser particuliers et sociétés et de les contacter directement, sans intermédiaire. La recherche est possible par lieu, par combinaison de langues et par spécialité.

De plus, depuis décembre 1997, des logiciels de traduction automatique sont en accès libre sur le web - à commencer par ceux de SYSTRAN sur AltaVista - et permettent de traduire en quelques secondes une page web ou un texte court, avec plusieurs combinaisons de langues possibles. Il va sans dire que la traduction automatique n'offre pas la qualité de travail des professionnels de la traduction, et qu'il est préférable de faire appel à ces derniers lorsqu'on a le temps et l'argent nécessaires. Ces logiciels sont toutefois très pratiques pour obtenir une traduction approximative en quelques secondes.

De plus en plus utilisée, la traduction assistée par ordinateur permet de coupler traduction automatique et travail du traducteur professionnel. Elle est une branche de l'ingénierie du langage, tout comme le traitement de la langue naturelle et la traduction automatique (traitée dans le chapitre suivant).

Lancé en janvier 1999 par la Commission européenne, le site HLTCentral (HLT: Human Languages Technologies) propose une courte définition de l'ingénierie du langage: «L'ingénierie du langage permet de vivre en toute convivialité avec la technologie. Nous pouvons utiliser notre connaissance du langage pour développer des systèmes capables de reconnaître à la fois la parole et l'écrit, de comprendre un texte suffisamment en profondeur pour être capable de sélectionner des informations, de le traduire dans différentes langues et de générer aussi bien un discours oral qu'un texte imprimé. L'application de ces technologies nous permet de repousser les limites actuelles de notre utilisation du langage. Les systèmes à commande vocale sont appelés à jouer un rôle prépondérant et à faire partie intégrante de notre vie quotidienne.»

Contrairement à la traduction automatique (TA) qui analyse le texte dans la langue source et génère automatiquement le texte correspondant dans la langue cible, sans intervention humaine pendant ce processus, la traduction assistée par ordinateur (TAO) est basée sur l'interaction entre l'homme et la machine pendant le processus de traduction.

La TAO est par exemple adoptée dès le milieu des années 1990 par le Bureau des services linguistiques de l'Organisation mondiale de la santé (OMS) à Genève (Suisse). Ce bureau travaille dans les six langues officielles de l'organisation: anglais, arabe, chinois, espagnol, français et russe.

Des expériences de traduction automatique (TA) sont également tentées, à plusieurs reprises, mais les traductions obtenues demandent un travail de révision trop important, si bien que, compte-tenu du type de documents à traduire, la TAO semble bien préférable.

Au sein de l'OMS, l'Unité de traduction assistée par ordinateur et de terminologie (CTT) explore les possibilités techniques offertes par les systèmes les plus récents de TAO, qui reposent sur la notion de «mémoire de traduction».

Comme expliqué sur le site web, «ces systèmes permettent au traducteur d'avoir immédiatement accès au patrimoine du "déjà traduit" dans lequel il peut puiser, quitte à rejeter ou modifier les solutions retenues par ses prédécesseurs, son choix définitif venant ensuite enrichir la mémoire. Ainsi, en archivant la production quotidienne, le traducteur aurait vite à sa disposition une "mémoire" colossale de solutions toutes faites à un nombre important de problèmes de traduction.»

En complément, le CTT utilise aussi plusieurs applications pour l'archivage électronique et la recherche en texte intégral, l'alignement de textes bilingues et multilingues, la gestion de mémoires de traduction et de bases de données terminologiques, et enfin la reconnaissance vocale.

Basé à Washington, D.C., au sein de l'Organisation panaméricaine de la santé (OPS), le Bureau régional de l'OMS pour les Amériques utilise par contre un système de traduction automatique développé par les linguistes computationnels, traducteurs et programmeurs de l'OPS.

Le service de traduction utilise SPANAM (de l'espagnol vers l'anglais) depuis 1980 et ENGSPAN (de l'anglais vers l'espagnol) depuis 1985, ce qui lui a permis de traiter plus de 25 millions de mots entre 1980 et 1998 dans les deux langues officielles de l'OPS. Le personnel et les traducteurs extérieurs post-éditent ensuite l'information brute avec un gain de productivité de 30 à 50%. Le système est installé sur le réseau local du siège de l'organisation et dans plusieurs bureaux régionaux pour pouvoir être utilisé par le personnel des services techniques et administratifs. Il est également diffusé auprès d'organismes publics et d'organismes à but non lucratif aux États-Unis, en Amérique latine et en Espagne. Ce système est plus tard renommé PAHOMTS, avec l'introduction de nouvelles paires de langues pour le portugais.

Autre expérience, celle de Wordfast. Lancé en 1999 par Yves Champollion, Wordfast est un logiciel de traduction avec terminologie disponible en temps réel et contrôle typographique. Il est compatible avec d'autres logiciels très utilisés comme le WebSphere Translation Server d’IBM et les logiciels de TMX ou de Trados. Une version simplifiée de Wordfast est téléchargeable gratuitement, avec un manuel d’utilisation disponible en seize langues.

Wordfast devient au fil des ans le numéro un mondial des logiciels de traduction utilisables sous toute plateforme (Windows, Mac, Linux), et le numéro deux mondial en nombre de ventes (après SDL Trados), avec 20.000 clients dans le monde en 2009, dont les Nations Unies, Nomura Securities, la NASA (National Aeronautics and Space Administration) et McGraw-Hill.

La traduction automatique

Définition
Historique
Quelques avis
Un marché en 2000
L'UNL, métalangage numérique

Définition

Un logiciel de traduction automatique (TA) analyse le texte dans la langue à traduire (langue source) et génère automatiquement le texte dans la langue désirée (langue cible), en utilisant des règles précises pour le transfert de la structure grammaticale. L'être humain n'intervient pas au cours du processus, contrairement à la traduction assistée par ordinateur (TAO), qui implique une interaction entre l'homme et la machine.

Si la traduction automatique reste très approximative, les logiciels de traduction sont toutefois très pratiques pour fournir un résultat immédiat et à moindres frais sinon gratuit. Ces logiciels n'ont cessé de s'améliorer au fil des ans, sans toutefois avoir la prétention d'égaler le travail du cerveau humain. De plus, depuis décembre 1997, des logiciels en accès libre sur le web permettent de traduire en quelques secondes une page web ou un texte court, avec plusieurs combinaisons de langues possibles.

SYSTRAN, société franco-américaine pionnière dans le traitement automatique des langues, explique sur son site web: «Un logiciel de traduction automatique traduit une langue naturelle dans une autre langue naturelle. La traduction automatique prend en compte la structure grammaticale de chaque langue et elle utilise des règles pour transférer la structure grammaticale de la langue source (texte à traduire) vers la langue cible (texte traduit). La traduction automatique ne remplace pas et n'est pas destinée à remplacer le traducteur humain.»

L'EAMT (European Association for Machine Translation) donne pour sa part la définition suivante: «La traduction automatique (TA) est l'utilisation de l'ordinateur pour la traduction de textes d'une langue naturelle à une autre. Elle fut un des premiers domaines de recherche en informatique. Il s'est avéré que cet objectif était difficile à atteindre. Cependant il existe aujourd'hui un certain nombre de systèmes produisant un résultat qui, s'il n'est pas parfait, est de qualité suffisante pour être utile dans certaines applications spécifiques, en général dans le domaine de la documentation technique. De plus, les logiciels de traduction, qui sont essentiellement destinés à aider le traducteur humain à produire des traductions, jouissent d'une popularité croissante auprès d'organisations de traduction professionnelles.»

Historique

L'intéressant historique donné sur le site de Globalink, une société spécialisée dans les logiciels et services de traduction (disparue depuis), est résumé dans les paragraphes suivants.

Dès leurs débuts, la traduction automatique et le traitement de la langue naturelle

progressent de pair avec l'évolution de l'informatique quantitative. Pendant la Seconde guerre mondiale, le développement des premiers ordinateurs programmables est accéléré par les premiers efforts cryptographiques pour tenter de fissurer les codes secrets allemands et autres codes de guerre. Suite à la guerre, la traduction et l'analyse du texte en langue naturelle procurent une base de travail au secteur émergent de la théorie de l'information.

Pendant les années 1950, la recherche sur la traduction automatique prend forme au sens de traduction littérale (mot à mot) sans utiliser de règles linguistiques. Le projet débuté à l'Université de Georgetown au début des années 1950 représente la première tentative systématique pour créer un système de traduction automatique utilisable pour le russe.

Jusqu'au milieu des années 1960, plusieurs projets de recherche aussi bien universitaires que gouvernementaux sont menés aux États-Unis et en Europe. Au même moment, les progrès rapides dans le domaine de la linguistique théorique culminent en 1965 avec la publication du livre *Aspects de la théorie syntaxique* de Noam Chomsky, et transforment radicalement les concepts permettant de comprendre la phonologie, la morphologie, la syntaxe et la sémantique du langage humain.

En 1966, le rapport ALPAC (Automatic Language Processing Advisory Committee) remis au gouvernement des États-Unis fait une estimation prématurément négative de la valeur des systèmes de traduction automatique et des perspectives sur leurs applications pratiques, mettant ainsi fin au financement et à l'expérimentation dans ce domaine pour la décennie suivante.

Il faut attendre la fin des années 1970 pour que des expériences sérieuses soient à nouveau entreprises, parallèlement aux progrès de l’informatique et des technologies des langues. Cette période voit le développement de systèmes de transfert d’une langue à l’autre et le lancement des premières tentatives commerciales. Des sociétés comme SYSTRAN et METAL sont persuadées de l'utilité et de la viabilité d’un tel marché. Elles mettent sur pied des produits et services de traduction automatique reliés à un serveur central.

Mais les problèmes restent nombreux, par exemple des coûts élevés de développement, un énorme travail lexicographique, la difficulté de proposer de nouvelles combinaisons de langues, l’inaccessibilité de tels systèmes pour l’utilisateur moyen, et enfin la difficulté de passer à de nouveaux stades de développement.

Quelques avis

Contrairement aux prévisions optimistes des années 1950 annonçant l'apparition imminente de la machine à traduire universelle, les systèmes de traduction automatique ne produisent pas encore de traductions de bonne qualité. Pourquoi? Pierre Isabelle et Patrick Andries, du laboratoire RALI (Laboratoire de recherche appliquée en linguistique informatique) à Montréal (Québec), expliquent ce échec dans *La traduction automatique, 50 ans après*, un article publié en 1998 dans les dossiers du magazine en ligne *Multimédium.*

Ils écrivent: «L'objectif ultime de construire une machine capable de rivaliser avec le

traducteur humain n'a cessé de fuir par devant les lentes avancées de la recherche. Les approches traditionnelles à base de règles ont conduit à des systèmes qui tendent à s'effondrer sous leur propre poids bien avant de s'élever au-dessus des nuages de l'ambiguïté sémantique. Les approches récentes à base de gros ensembles de textes, appelés corpus - qu'elles soient fondées sur les méthodes statistiques ou les méthodes analogiques - promettent bien de réduire la quantité de travail manuel requise pour construire un système de TA [traduction automatique], mais il est moins sûr qu'elles promettent des améliorations substantielles de la qualité des traductions machine.»

Reprenant les idées de Yehochua Bar-Hillel exprimées dans *The State of Machine Translation*, un article publié en 1951, Pierre Isabelle et Patrick Andries définissent trois stratégies d'application de la traduction automatique: (a) une aide pour «balayer» la production écrite et fournir des traductions approximatives; (b) des situations de «sous-langues naturelles simples», comme l'implantation réussie en 1977 du système METEO qui traduit les prévisions météorologiques du ministère de l'Environnement canadien; (c) et enfin, pour de bonnes traductions de textes complexes, le couplage de l'humain et de la machine avant, pendant et après le processus de traduction automatique, un couplage qui n'est pas forcément économique comparé à la traduction traditionnelle.

Les auteurs penchent plus pour «un poste de travail pour le traducteur humain» que pour un «traducteur robot». Ils expliquent: «Les recherches récentes sur les méthodes probabilistes ont permis de démontrer qu'il était possible de modéliser d'une manière extrêmement efficace certains aspects simples du rapport traductionnel entre deux textes. Par exemple, on a mis au point des méthodes qui permettent de calculer le bon "appariement" entre les phrases d'un texte et de sa traduction, c'est-à-dire d'identifier à quelle(s) phrase(s) du texte d'origine correspond chaque phrase de la traduction. Appliquées à grande échelle, ces techniques permettent de constituer, à partir des archives d'un service de traduction, un mémoire de traduction qui permettra souvent de recycler des fragments de traductions antérieures. Des systèmes de ce genre ont déjà commencé à apparaître sur le marché (Translation Manager II de IBM, Translator's Workbench de Trados, TransSearch du RALI, etc.). Les recherches les plus récentes se concentrent sur des modèles capables d'établir automatiquement les correspondances à un niveau plus fin que celui de la phrase: syntagmes et mots. Les résultats obtenus laissent entrevoir toute une famille de nouveaux outils pour le traducteur humain, dont les aides au dépouillement terminologique, les aides à la dictée et à la frappe des traductions ainsi que les détecteurs de fautes de traduction.»

Fondé en 1990, Globalink est une société spécialisée dans les logiciels et services de traduction. Elle offre des solutions sur mesure à partir d'une gamme de logiciels, options en ligne et services de traduction professionnelle. La société diffuse ses logiciels de traduction en allemand, anglais, espagnol, français, italien et portugais, et propose des solutions aux problèmes de traduction des particuliers, petites sociétés, multinationales et gouvernements, que ce soit pour un produit individuel donnant une traduction préliminaire rapide ou un système complet permettant de gérer des traductions de documents professionnels.

Le site web donne les informations suivantes en 1998: «Avec les logiciels d'application de Globalink, l'ordinateur utilise trois ensembles de données: le texte à traiter, le programme de traduction et un dictionnaire de mots et d'expressions dans la langue-source, ainsi que des informations sur les concepts évoqués par le dictionnaire et les

règles applicables à la phrase: règles de syntaxe et de grammaire, y compris des algorithmes gouvernant la conjugaison des verbes, l'adaptation de la syntaxe, les accords de genre et de nombre et la mise en ordre des mots. Une fois que l'utilisateur a sélectionné le texte et lancé le processus de traduction, le programme commence à comparer les mots du texte à traiter avec ceux qui sont stockés dans le dictionnaire. Une fois l'adéquation trouvée, l'application prépare une notice complète qui inclut des informations sur les significations possibles du mot et, d'après le contexte, ses relations avec les autres mots dans la même phrase. Le temps requis pour la traduction dépend de la longueur du texte. Un document de trois pages et 750 mots demande un traitement de trois minutes environ pour une première traduction.»

En septembre 1998, Randy Hobler, consultant en marketing internet auprès de Globalink, est résolument optimiste: «Nous arriverons rapidement au point où une traduction très fidèle du texte et de la parole sera si commune qu'elle pourra faire partie des plateformes ou même des puces. A ce stade, lorsque le développement de l'internet aura atteint sa vitesse de croisière, lorsque la fidélité de la traduction atteindra plus de 98% et lorsque les différentes combinaisons de langues possibles auront couvert la grande majorité du marché, la transparence de la langue - à savoir toute communication d'une langue à une autre - sera une vision trop restrictive pour ceux qui vendent cette technologie. Le développement suivant sera la "transparence transculturelle et transnationale" dans laquelle les autres aspects de la communication humaine, du commerce et des transactions au-delà du seul langage entreront en scène. Par exemple, les gestes ont un sens, les mouvements faciaux ont un sens, et ceci varie en fonction des normes sociales d'un pays à l'autre. La lettre O réalisée avec le pouce et l'index signifie "OK" aux États-Unis alors qu'en Argentine c'est un geste obscène.

Quand se produira l'inévitable développement de la vidéoconférence multilingue multimédia, il sera nécessaire de corriger visuellement les gestes. Le Media Lab du MIT [Massachusetts Institute of Technology], Microsoft et bien d'autres travaillent à la reconnaissance informatique des expressions faciales, l'identification des caractéristiques biométriques par le biais du visage, etc. Il ne servira à rien à un homme d'affaires américain de faire une excellente présentation à un Argentin lors d'une vidéoconférence multilingue sur le web, avec son discours traduit dans un espagnol argentin parfait, s'il fait en même temps le geste O avec le pouce et l'index. Les ordinateurs pourront intercepter ces gestes et les corriger visuellement. Les cultures diffèrent de milliers de façons, et la plupart de leurs codes peuvent être modifiés par voie informatique lorsqu'on passe de l'un à l'autre. Ceci inclut les lois, les coutumes, les habitudes de travail, l'éthique, le change monétaire, les différences de taille dans les vêtements, les différences entre le système métrique et le système de mesure anglophone, etc. Les firmes dynamiques répertorieront et programmeront ces différences, et elles vendront des produits et services afin d'aider les habitants de la planète à mieux communiquer entre eux. Une fois que ces produits et services seront largement répandus, ils contribueront réellement à une meilleure compréhension à l'échelle internationale.»

D'autres sont moins enthousiastes.

François Vadrot, directeur de FTPress (French Touch Press), société de presse en ligne, écrit en mai 2000: «Peut-on réellement penser que toute la population du monde va communiquer dans tous les sens? Peut-être? Via des systèmes de traduction

instantanée, par écrit ou par oral? J'ai du mal à imaginer qu'on verra de sitôt des outils capables de translater les subtilités des modes de pensée propres à un pays: il faudrait pour lors traduire, non plus du langage, mais établir des passerelles de sensibilité. A moins que la mondialisation n'uniformise tout cela?»

Alex Andrachmes, producteur audiovisuel, écrivain et explorateur d'hypertexte, est tout aussi dubitatif. Il écrit en décembre 2000: «J'attends les fameuses traductions simultanées en direct-live... On nous les annonce avec les nouveaux processeurs ultra-puissants, mais on nous les annonçait déjà pour cette génération-ci de processeurs. Alors, le genre: vous/réservé/avion/de le/november 17-2000... Non merci. Plus tard peut-être.»

Les progrès sont toutefois rapides et permettent de rêver à des applications simples et universelles. Tel est le souhait de Tim McKenna, écrivain et philosophe, qui écrit en octobre 2000: «Quand la qualité des logiciels sera suffisante pour que les gens puissent converser sur le web par écrit ou par oral en temps réel dans différentes langues, nous verrons tout un monde s'ouvrir à nous. Les scientifiques, les hommes politiques, les hommes d'affaires et bien d'autres groupes seront à même de communiquer immédiatement entre eux sans l'intermédiaire de médiateurs ou traducteurs.»

Un marché en 2000

En 2000, la généralisation de l'internet et du commerce électronique entraîne le développement d'un véritable marché, avec les produits et services des sociétés SYSTRAN, Alis Technologies, Lernout & Hauspie, Globalink et Softissimo, etc., avec des gammes de produits ciblant le grand public, les professionnels et les industriels.

SYSTRAN (acronyme de: System Translation), pionnier dans le traitement automatique des langues, est l'auteur du premier logiciel de traduction gratuit du web, lancé en décembre 1997. AltaVista Translation, appelé aussi Babel Fish, traduit les pages web de l'anglais vers les langues suivantes : allemand, français, espagnol, italien et portugais, et vice versa. Le siège de SYSTRAN est situé à Soisy-sous-Montmorency (France). Sa succursale, située à La Jolla (Californie), assure les ventes et le marketing, ainsi qu'une partie de la R&D (recherche et développement).

Basée à Montréal (Québec), Alis Technologies développe et commercialise des solutions et services de traitement linguistique au moyen de logiciels de traduction qui transforment des systèmes informatiques unilingues en outils multilingues.

Basé à Ypres (Belgique) et Burlington (Massachusetts, États-Unis), Lernout & Hauspie (racheté ensuite par ScanSoft) propose des produits et services en dictée, traduction, compression vocale, synthèse vocale et documentation industrielle. Les technologies couvertes incluent la reconnaissance automatique de la langue, la compression numérique de la parole, le passage du texte à la parole, et la traduction. Les produits émanant des trois premières technologies sont vendus aux grandes sociétés des industries suivantes: télécommunications, informatique, multimédia, électronique grand public et électronique automotrice. Les services de traduction sont à destination des sociétés en technologies de l'information, des marchés verticaux et des marchés d'automatisation.

De plus, le Machine Translation Group formé par Lernout & Hauspie comprend des entreprises qui développent, produisent et vendent des systèmes de traduction: L&H Language Technology, AppTek, AILogic, NeocorTech et Globalink.

La société Softissimo commercialise la série de logiciels de traduction Reverso, à côté de produits d’écriture multilingue, de dictionnaires électroniques et de méthodes de langues. Reverso est utilisé par exemple par Voilà, le moteur de recherche de France Télécom. Softissimo diffuse aussi des logiciels d'apprentissage des langues, ainsi que des dictionnaires, notamment l'Eurodico, le *Grand Collins bilingue* et le *Collins English Dictionary*.

En mars 2001, IBM se lance dans un marché en pleine expansion avec un produit professionnel haut de gamme, le WebSphere Translation Server. Ce logiciel traduit instantanément en plusieurs langues (allemand, anglais, chinois, coréen, espagnol, français, italien, japonais) les pages web, courriels et chats (dialogues en direct). Il interprète 500 mots à la seconde et permet l’ajout de vocabulaires spécifiques.

L'UNL, métalangage numérique

Développé à Tokyo (Japon) sous l'égide de l'Institute of Advanced Studies (IAS) de l'Université des Nations Unies (UNU), l'UNL (universal networking language) est un projet de métalangage numérique pour l'encodage, le stockage, la recherche et la communication d'informations multilingues. Il s'agirait d'une interlangue formant une passerelle entre une langue source et une langue cible et offrant ainsi une solution au problème de communication posé par la barrière des langues.

Ce métalangage est développé à partir de janvier 1997 au sein de l'UNL Program, un programme international impliquant de nombreux partenaires dans toutes les communautés linguistiques. En 1998, 120 chercheurs de par le monde travaillent sur un projet multilingue comportant seize langues (allemand, anglais, arabe, brésilien, chinois, espagnol, français, hindou, indonésien, italien, japonais, letton, mongolien, russe, swahili et thaï).

Ce programme se poursuit ensuite sous l'égide de la Fondation UNDL (Universal Networking Digital Language), créée en janvier 2001 pour développer et promouvoir le projet UNL, avec un siège social à Genève (Suisse). En 2004, la Fondation UNDL est accréditée par les Nations Unies en tant qu'organisation non gouvernementale (ONG).

Comme expliqué dans un des wikis du projet en 2010, «l'UNL est une langue artificielle créée pour prendre en compte les informations et connaissances véhiculées par les langues humaines. Elle est dotée de composantes lexicales, grammaticales et sémantiques, comme les langues naturelles. Couplée à l’intelligence artificielle, l’UNL facilite la communication entre l’homme et la machine, et par le biais de la machine, entre tous les peuples dans la langue maternelle de chacun. Notre première tâche est de compléter le système UNL. Ensuite, le mettre au service des toutes les nations.»

Chronologie

Cette chronologie ne prétend pas à l'exhaustivité. Subjective, elle vise seulement à souligner quelques grandes étapes. Chaque ligne débute par l'année ou bien l'année/mois. Par exemple, 1971/07 signifie juillet 1971.

1968: Le code ASCII est le premier système d'encodage informatique.
1971/07: Le Projet Gutenberg est la première bibliothèque numérique.
1974: L'internet fait ses débuts.
1990: Le web est inventé par Tim Berners-Lee.
1991/01: L'Unicode est un système d'encodage pour toutes les langues.
1993/11: Mosaic est le premier logiciel de navigation sur le web.
1994: Travlang est un site consacré aux langues et aux voyages.
1994/04: La Human-Languages Page est un catalogue de ressources linguistiques.
1994/10: Le World Wide Web Consortium offre des outils pour un web multilingue.
1995: L'Internet Dictionary Project lance des dictionnaires de traduction gratuits.
1995: NetGlos est un glossaire multilingue pour la terminologie de l'internet.
1995: Global Reach est une société de conseil pour la localisation des sites.
1995: L'Ethnologue est disponible gratuitement sur le web.
1995/12: La Kotoba Home Page explique comment afficher plusieurs langues.
1996/04: OneLook Dictionaries offre un point commun pour consulter des dictionnaires.
1997/01: L'UNL (universal networking language) est un projet de métalangage numérique.
1997: Le Logos Dictionary est mis en ligne gratuitement.
1997/12: AltaVista lance AltaVista Translation, appelé aussi Babel Fish.
1999: WordReference.com propose des dictionnaires bilingues gratuits.
1999: Wordfast propose son premier logiciel de traduction.
1999/12: WebEncyclo est la première grande encyclopédie francophone en ligne.
1999/12: Britannica.com est la première grande encyclopédie anglophone en ligne.
2000/02: yourDictionary.com est un portail pour les dictionnaires.
2000/07: La moitié des usagers de l'internet est non anglophone.
2000/09: Le Grand dictionnaire terminologique (GDT) est bilingue français-anglais.
2001/01: Wikipédia est la première grande encyclopédie collaborative multilingue.
2001/01: L'UNL est désormais développé au sein de la Fondation UNDL.
2001/04: La Human-Languages Page devient le portail iLoveLanguages.
2004/01: Le Projet Gutenberg Europe propose une interface multilingue.
2007/03: IATE est la base terminologique multilingue européenne.
2009: La nouvelle édition (16e) de l'Ethnologue recense 6.909 langues.

Remerciements

Ce livre doit beaucoup à toutes les personnes ayant accepté de répondre à mes questions au fil des ans. Certains entretiens sont disponibles en ligne sur le NEF (Net des études françaises), Université de Toronto <www.etudes-francaises.net/entretiens/>. D'autres entretiens ont été directement inclus dans ce livre ou alors ils ont inspiré des idées développées dans ces pages.

Merci à Nicolas Ancion, Alex Andrachmes, Guy Antoine, Silvaine Arabo, Arlette Attali, Marc Autret, Isabelle Aveline, Jean-Pierre Balpe, Emmanuel Barthe, Robert Beard, Michael Behrens, Michel Benoît, Guy Bertrand, Olivier Bogros, Christian Boitet, Bernard Boudic, Bakayoko Bourahima, Marie-Aude Bourson, Lucie de Boutiny, Anne-Cécile Brandenbourger, Alain Bron, Patrice Cailleaud, Tyler Chambers, Pascal Chartier, Richard Chotin, Alain Clavet, Jean-Pierre Cloutier, Jacques Coubard, Luc Dall'Armellina, Kushal Dave, Cynthia Delisle, Émilie Devriendt, Bruno Didier, Catherine Domain, Helen Dry, Bill Dunlap, Pierre-Noël Favennec, Gérard Fourestier, Pierre François Gagnon, Olivier Gainon, Jacques Gauchey, Raymond Godefroy, Muriel Goiran, Marcel Grangier, Barbara Grimes, Michael Hart, Michael Kellogg, Roberto Hernández Montoya, Randy Hobler, Eduard Hovy, Christiane Jadelot, Gérard Jean-François, Jean-Paul, Anne-Bénédicte Joly, Michael Kellogg, Brian King, Geoffrey Kingscott, Steven Krauwer, Gaëlle Lacaze, Michel Landaret, Hélène Larroche, Pierre Le Loarer, Claire Le Parco, Annie Le Saux, Fabrice Lhomme, Philippe Loubière, Pierre Magnenat, Xavier Malbreil, Alain Marchiset, Maria Victoria Marinetti, Michael Martin, Tim McKenna, Emmanuel Ménard, Yoshi Mikami, Jacky Minier, Jean-Philippe Mouton, John Mark Ockerbloom, Caoimhín Ó Donnaíle, Jacques Pataillot, Alain Patez, Nicolas Pewny, Marie-Joseph Pierre, Hervé Ponsot, Olivier Pujol, Anissa Rachef, Peter Raggett, Patrick Rebollar, Philippe Renaut, Jean-Baptiste Rey, Philippe Rivière, Blaise Rosnay, Bruno de Sa Moreira, Pierre Schweitzer, Henk Slettenhaar, Murray Suid, June Thompson, Zina Tucsnak, François Vadrot, Christian Vandendorpe, Robert Ware, Russon Wooldridge et Denis Zwirn.

Section 2. Information about the Mission of Project Gutenberg-tm

Project Gutenberg-tm is synonymous with the free distribution of electronic works in formats readable by the widest variety of computers including obsolete, old, middle-aged and new computers. It exists because of the efforts of hundreds of volunteers and donations from people in all walks of life.

Volunteers and financial support to provide volunteers with the assistance they need are critical to reaching Project Gutenberg-tm's goals and ensuring that the Project Gutenberg-tm collection will remain freely available for generations to come. In 2001, the Project Gutenberg Literary Archive Foundation was created to provide a secure and permanent future for Project Gutenberg-tm and future generations. To learn more about the Project Gutenberg Literary Archive Foundation and how your efforts and donations can help, see Sections 3 and 4 and the Foundation web page at http://www.pglaf.org.

Section 3. Information about the Project Gutenberg Literary Archive Foundation

The Project Gutenberg Literary Archive Foundation is a non profit 501(c)(3) educational corporation organized under the laws of the state of Mississippi and granted tax exempt status by the Internal Revenue Service. The Foundation's EIN or federal tax identification number is 64-6221541. Its 501(c)(3) letter is posted at http://pglaf.org/fundraising. Contributions to the Project Gutenberg Literary Archive Foundation are tax deductible to the full extent permitted by U.S. federal laws and your state's laws.

The Foundation's principal office is located at 4557 Melan Dr. S. Fairbanks, AK, 99712., but its volunteers and employees are scattered throughout numerous locations. Its business office is located at 809 North 1500 West, Salt Lake City, UT 84116, (801) 596-1887, email business@pglaf.org. Email contact links and up to date contact information can be found at the Foundation's web site and official page at http://pglaf.org

For additional contact information:
Dr. Gregory B. Newby
Chief Executive and Director
gbnewby@pglaf.org

Section 4. Information about Donations to the Project Gutenberg Literary Archive Foundation

Project Gutenberg-tm depends upon and cannot survive without wide spread public support and donations to carry out its mission of increasing the number of public domain and licensed works that can be freely distributed in machine readable form accessible by the widest array of equipment including outdated equipment. Many small donations ($1 to $5,000) are particularly important to maintaining tax exempt status with the IRS.

The Foundation is committed to complying with the laws regulating charities and charitable donations in all 50 states of the United States. Compliance requirements are not uniform and it takes a considerable effort, much paperwork and many fees to meet and keep up with these requirements. We do not solicit donations in locations where we have not received written confirmation of compliance. To SEND DONATIONS or determine the status of compliance for any particular state visit http://pglaf.org